



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C07H 21/04, C12N 15/63, 15/85, 15/09, C07K 5/00, 14/00, C12P 21/00	A1	(11) International Publication Number: WO 00/55174 (43) International Publication Date: 21 September 2000 (21.09.00)
(21) International Application Number: PCT/US00/05988 (22) International Filing Date: 8 March 2000 (08.03.00) (30) Priority Data: 60/124,270 12 March 1999 (12.03.99) US (71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US). (71)(72) Applicant and Inventor: ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). (72) Inventor; and (75) Inventor/Applicant (for US only): RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Laytonsville, MD 20882 (US). (74) Agents: WALES, Michele, M. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i>
(54) Title: HUMAN PROSTATE CANCER ASSOCIATED GENE SEQUENCES AND POLYPEPTIDES (57) Abstract This invention relates to newly identified prostate or prostate cancer related polynucleotides and the polypeptides encoded by these polynucleotides herein collectively known as "prostate cancer antigens", and to the complete gene sequences associated therewith and to the expression products thereof, as well as the use of such prostate cancer antigens for detection, prevention and treatment of disorders of the prostate, particularly the presence of prostate cancer. This invention relates to the prostate cancer antigens as well as vectors, host cells, antibodies directed to prostate cancer antigens and recombinant and synthetic methods for producing the same. Also provided are diagnostic methods for diagnosing and treating, preventing and/or prognosing disorders related to the prostate, including prostate cancer, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying agonists and antagonists of prostate cancer antigens of the invention. The present invention further relates to methods and/or compositions for inhibiting the production and/or function of the polypeptides of the present invention.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Human Prostate Cancer Associated Gene Sequences and Polypeptides

5 *Field of the Invention*

This invention relates to newly identified prostate or prostate cancer related polynucleotides and the polypeptides encoded by these polynucleotides herein collectively known as "prostate cancer antigens," and to the complete gene sequences associated therewith and to the expression products thereof, as well as the use of such prostate cancer antigens for detection, prevention and treatment of disorders of the prostate, particularly the presense of prostate cancer. This invention relates to the prostate cancer antigens as well as vectors, host cells, antibodies directed to prostate cancer antigens and recombinant and synthetic methods for producing the same. Also provided are diagnostic methods for diagnosing and treating, preventing and/or prognosing disorders related to the prostate, including prostate cancer, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying agonists and antagonists of prostate cancer antigens of the invention. The present invention further relates to methods and/or compositions for inhibiting the production and/or function of the polypeptides of the present invention.

20

Background of the Invention

Cell growth is a carefully regulated process which responds to specific needs of the body. Occassionally, the intricate, and highly regulated controls dictating the rules for cellular division break down. When this occurs, the cell begins to grow and divide independently of its homeostatic regulation resulting in a condition commonly referred to as cancer. In fact, cancer is the second leading cause of death among Americans aged 25-44.

Prostate cancer has become the most common cancer among American men, and only lung cancer is responsible for more cancer deaths (Boring, Cancer Statistics, 41:19-36 (1991)). The age specific mortality rate has slowly increased over the past 50 years and in black American men is nearly double the rate found in white men (Carter, Prostate,

30

16:39-48 (1990)). Prostate cancer is responsible for nearly three percent of all deaths in men over the age of 55 years (Seidman, et al., Probabilities of Eventually Developing or Dying of Cancer-United States, 35:36-56 (1985)). Since the incidence of prostate cancer increases more rapidly with age than any other cancer, and the average age of American men is rising, the number of patients with prostate cancer is expected to increase dramatically over the next decade.

Approximately 30% of men with prostate cancer have distant metastases at the time of diagnosis (Schmidt, et al., J. Urol., 136:416-421 (1986)). Despite the impressive symptomatic response of metastases to hormonal manipulation (androgen deprivation), the survival rate for these patients is dismal: the median duration of survival is less than three years (Eyar, Urologic Pathology: The Prostate, Philadelphia, Pa., Lea and Febiger, 241-267 (1977)). By five years, over 75% and by ten years, more than 90% of these patients die of their cancer rather than with it (Silverberg, Cancer, 60:692-717 (1987) (Suppl.)). The problem with prostate cancer is that many forms of prostate cancer are latent, in other words, such forms are difficult to detect. Approximately 30% of the men over the age of 50 years who have no clinical evidence of prostate cancer harbor foci of cancer within the prostate (McNeal, et al., The Lancet, January, 11:60-63 (1986)). This remarkably high prevalence of prostate cancer at autopsy, seen in no other organ, makes it the most common malignancy in human beings (Dhom, J. Cancer Res. Clin. Oncol., 106:210-218 (1983)). There is strong support for the concept of multi-step process in the pathogenesis of prostate cancer in which latent cancers progress through some but not all of the steps necessary for full malignant expression (Utter, et al., J. Urol., 143:742-746 (1990)).

There are a variety of techniques for early detection and characteristics of prostate cancers, however, none of them are devoid of problems. Prostate cancer is a notoriously silent disease with few early symptoms. There is a need, therefore, for identification and characterization of factors that modulate activation and differentiation of prostate cells, both normally and in disease states. In particular, there is a need to isolate and characterize additional molecules that mediate apoptosis, DNA repair, tumor-mediated angiogenesis, genetic imprinting, immune responses to tumors and tumor antigens and, among other things, that can play a role in detecting, preventing, ameliorating or correcting dysfunctions or diseases related to the prostate.

Summary of the Invention

The present invention includes isolated nucleic acid molecules comprising, or alternatively, consisting of, a prostate and/or prostate cancer associated polynucleotide sequence disclosed in the sequence listing (as SEQ ID Nos:1 to 940) and/or contained in a human cDNA clone described in Tables 1, 2 and 5 and deposited with the American Type Culture Collection ("ATCC"). Fragments, variant, and derivatives of these nucleic acid molecules are also encompassed by the invention. The present invention also includes isolated nucleic acid molecules comprising, or alternatively consisting of, a polynucleotide encoding a prostate or prostate cancer polypeptide. The present invention further includes prostate and/or prostate cancer polypeptides encoded by these polynucleotides. Further provided for are amino acid sequences comprising, or alternatively consisting of, prostate and/or prostate cancer polypeptides as disclosed in the sequence listing (as SEQ ID Nos: 941 to 1880) and/or encoded by a human cDNA clone described in Tables 1, 2 and 5 and deposited with the ATCC. Antibodies that bind these polypeptides are also encompassed by the invention. Polypeptide fragments, variants, and derivatives of these amino acid sequences are also encompassed by the invention, as are polynucleotides encoding these polypeptides and antibodies that bind these polypeptides. Also provided are diagnostic methods for diagnosing and treating, preventing, and/or prognosing disorders related to the prostate, including prostate cancer, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying agonists and antagonists of prostate cancer antigens of the invention.

Detailed Description

Tables

Table 1 summarizes some of the prostate cancer antigens encompassed by the invention (including contig sequences (SEQ ID NO:X) and the cDNA clone related to the contig sequence) and further summarizes certain characteristics of the prostate cancer polynucleotides and the polypeptides encoded thereby. The first column shows the "SEQ ID NO:" for each of the 940 prostate cancer antigen polynucleotide sequences of the invention. The second column provides a unique "Sequence/Contig ID" identification for

each prostate and/or prostate cancer associated sequence. The third column, "Gene Name," and the fourth column, "Overlap," provide a putative identification of the gene based on the sequence similarity of its translation product to an amino acid sequence found in a publicly accessible gene database and the database accession no. for the database sequence having similarity, respectively. The fifth and sixth columns provide the location (nucleotide position nos. within the contig), "Start" and "End", in the polynucleotide sequence "SEQ ID NO:X" that delineate the preferred ORF shown in the sequence listing as SEQ ID NO:Y. The seventh and eighth columns provide the "% Identity" (percent identity) and "% Similarity" (percent similarity), respectively, observed between the aligned sequence segments of the translation product of SEQ ID NO:X and the database sequence. The ninth column provides a unique "Clone ID" for a cDNA clone related to each contig sequence.

Table 2 summarizes ATCC Deposits, Deposit dates, and ATCC designation numbers of deposits made with the ATCC in connection with the present application.

Table 3 indicates public ESTs, of which at least one, two, three, four, five, ten, fifteen or more of any one or more of these public EST sequences are optionally excluded from certain embodiments of the invention.

Table 4 lists residues comprising antigenic epitopes of antigenic epitope-bearing fragments present in most of the prostate or prostate cancer associated polynucleotides described in Table 1 as predicted by the inventors using the algorithm of Jameson and Wolf, (1988) Comp. Appl. Biosci. 4:181-186. The Jameson-Wolf antigenic analysis was performed using the computer program PROTEAN (Version 3.11 for the Power MacIntosh, DNASTAR, Inc., 1228 South Park Street Madison, WI). Prostate and prostate cancer associated polypeptides (e.g., SEQ ID NO:Y, polypeptides encoded by SEQ ID NO:X, or polypeptides encoded by the cDNA in the referenced cDNA clone) may possess one or more antigenic epitopes comprising residues described in Table 4. It will be appreciated that depending on the analytical criteria used to predict antigenic determinants, the exact address of the determinant may vary slightly. The residues and locations shown in column two of Table 4 correspond to the amino acid sequences for most prostate and prostate cancer associated polypeptide sequence shown in the Sequence Listing.

Table 5 shows the cDNA libraries sequenced, and ATCC designation numbers

and vector information relating to these cDNA libraries.

Definitions

5 The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

 In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide
10 could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide. The term "isolated" does not refer to genomic or cDNA libraries, whole cell total or mRNA preparations, genomic DNA preparations (including those separated by electrophoresis and transferred onto blots),
15 sheared whole cell genomic DNA preparations or other compositions where the art demonstrates no distinguishing features of the polynucleotide/sequences of the present invention.

 As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X (as described in column 1 of Table 1) or the related
20 cDNA clone (as described in column 9 of Table 1 and contained within a library deposited with the ATCC). For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule
25 having an amino acid sequence encoded by a polynucleotide of the invention as broadly defined (obviously excluding poly-Phenylalanine or poly-Lysine peptide sequences which result from translation of a polyA tail of a sequence corresponding to a cDNA).

 In the present invention, "SEQ ID NO:X" was often generated by overlapping sequences contained in multiple clones (contig analysis). A representative clone
30 containing all or most of the sequence for SEQ ID NO:X is deposited at Human Genome Sciences, Inc. (HGS) in a catalogued and archived library. As shown in column 9 of Table 1, each clone is identified by a cDNA Clone ID. Each Clone ID is unique to an

individual clone and the Clone ID is all the information needed to retrieve a given clone from the HGS library. In addition to the individual cDNA clone deposits, most of the cDNA libraries from which the clones were derived were deposited at the American Type Culture Collection (hereinafter "ATCC"). Table 5 provides a list of the deposited cDNA libraries. One can use the Clone ID to determine the library source by reference to Tables 2 and 5. Table 5 lists the deposited cDNA libraries by name and links each library to an ATCC Deposit. Library names contain four characters, for example, "HTWE." The name of a cDNA clone ("Clone ID") isolated from that library begins with the same four characters, for example "HTWEP07". As mentioned below, Table 1 correlates the Clone ID names with SEQ ID NOs. Thus, starting with a SEQ ID NO, one can use Tables 1, 2 and 5 to determine the corresponding Clone ID, from which library it came and in which ATCC deposit the library is contained. Furthermore, it is possible to retrieve a given cDNA clone from the source library by techniques known in the art and described elsewhere herein. The ATCC is located at 10801 University Boulevard, Manassas, Virginia 20110-2209, USA. The ATCC deposits were made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for the purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, or the complement thereof (e.g., the complement of any one, two, three, four, or more of the polynucleotide fragments described herein), and/or sequences contained in the related cDNA clone within a library deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42 degree C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65 degree C.

Also included within "polynucleotides" of the present invention are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt

conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37 degree C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH_2PO_4 ; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA: followed by washes at 50 degree C with 1XSSPE, 0.1%
5 SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent,
10 BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as
15 any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone generated using oligo dT as a primer).

20 The polynucleotides of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-
25 stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other
30 reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

In specific embodiments, the polynucleotides of the invention are at least 15, at least 30, at least 50, at least 100, at least 125, at least 500, or at least 1000 continuous nucleotides but are less than or equal to 300 kb, 200 kb, 100 kb, 50 kb, 15 kb, 10 kb, 7.5kb, 5 kb, 2.5 kb, 2.0 kb, or 1 kb, in length. In a further embodiment, polynucleotides of the invention comprise a portion of the coding sequences, as disclosed herein, but do not comprise all or a portion of any intron. In another embodiment, the polynucleotides comprising coding sequences do not contain coding sequences of a genomic flanking gene (i.e., 5' or 3' to the gene of interest in the genome). In other embodiments, the polynucleotides of the invention do not contain the coding sequence of more than 1000, 500, 250, 100, 50, 25, 20, 15, 10, 5, 4, 3, 2, or 1 genomic flanking gene(s).

"SEQ ID NO:X" refers to a prostate cancer antigen polynucleotide sequence described in Table 1. SEQ ID NO:X is identified by an integer specified in column 1 of Table 1. The polypeptide sequence SEQ ID NO:Y is a translated open reading frame (ORF) encoded by polynucleotide SEQ ID NO:X. There are 940 prostate cancer antigen polynucleotide sequences described in Table 1 and shown in the sequence listing (SEQ ID NO:1 through SEQ ID NO:940). Likewise there are 940 polypeptide sequences shown in the sequence listing, one polypeptide sequence for each of the polynucleotide sequences (SEQ ID NO:941 through SEQ ID NO:1880). The polynucleotide sequences are shown in the sequence listing immediately followed by all of the polypeptide sequences. Thus, a polypeptide sequence corresponding to polynucleotide sequence SEQ ID NO:1 is the first polypeptide sequence shown in the sequence listing. The second polypeptide sequence corresponds to the polynucleotide sequence shown as SEQ ID NO:2, and so on. In other words, since there are 940 polynucleotide sequences, for any polynucleotide sequence SEQ ID NO:X, a corresponding polypeptide SEQ ID NO:Y can be determined by the formula $X + 940 = Y$. In addition, any of the unique "Sequence/Contig ID" defined in column 2 of Table 1, can be linked to the corresponding polypeptide SEQ ID NO:Y by reference to Table 4.

The polypeptides of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications

are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

The prostate and prostate cancer polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below). It is often advantageous to include an additional amino acid sequence which contains secretory or

leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The prostate and prostate cancer polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified using techniques described herein or otherwise known in the art, such as, for example, by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988). Polypeptides of the invention also can be purified from natural, synthetic or recombinant sources using techniques described herein or otherwise known in the art, such as, for example, antibodies of the invention raised against the polypeptides of the present invention in methods which are well known in the art.

By a polypeptide demonstrating a "functional activity" is meant, a polypeptide capable of displaying one or more known functional activities associated with a full-length (complete) protein of the invention. Such functional activities include, but are not limited to, biological activity, antigenicity [ability to bind (or compete with a polypeptide for binding) to an anti-polypeptide antibody], immunogenicity (ability to generate antibody which binds to a specific polypeptide of the invention), ability to form multimers with polypeptides of the invention, and ability to bind to a receptor or ligand for a polypeptide.

"A polypeptide having functional activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular assay, such as, for example, a biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention).

The functional activity of the prostate cancer antigen polypeptides, and fragments, variants derivatives, and analogs thereof, can be assayed by various methods.

For example, in one embodiment where one is assaying for the ability to bind or compete with full-length polypeptide of the present invention for binding to an antibody to the full length polypeptide antibody, various immunoassays known in the art can be used, including but not limited to, competitive and non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitation reactions, immunodiffusion assays, in situ immunoassays (using colloidal gold, enzyme or radioisotope labels, for example), western blots, precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention.

In another embodiment, where a ligand is identified, or the ability of a polypeptide fragment, variant or derivative of the invention to multimerize is being evaluated, binding can be assayed, e.g., by means well-known in the art, such as, for example, reducing and non-reducing gel chromatography, protein affinity chromatography, and affinity blotting. See generally, Phizicky, E., et al., Microbiol. Rev. 59:94-123 (1995). In another embodiment, physiological correlates polypeptide of the present invention binding to its substrates (signal transduction) can be assayed.

In addition, assays described herein (see Examples) and otherwise known in the art may routinely be applied to measure the ability of polypeptides of the present invention and fragments, variants derivatives and analogs thereof to elicit polypeptide related biological activity (either in vitro or in vivo). Other methods will be known to the skilled artisan and are within the scope of the invention.

Prostate and Prostate Cancer Associated Polynucleotides and Polypeptides of the Invention

It has been discovered herein that the polynucleotides described in Table 1 are expressed at significantly enhanced levels in human prostate and/or prostate cancer tissues. Accordingly, such polynucleotides, polypeptides encoded by such polynucleotides, and antibodies specific for such polypeptides find use in the prediction, diagnosis, prevention and treatment of prostate related disorders, including prostate cancer as more fully described below.

Table 1 summarizes some of the polynucleotides encompassed by the invention (including contig sequences (SEQ ID NO:X) and the related cDNA clones) and further summarizes certain characteristics of these prostate and/or prostate cancer associated polynucleotides and the polypeptides encoded thereby.

13

Table 1

Seq ID No.	Sequence/ Contig ID	Gene Name	Overlap	HGS Nucleotide Start End	% Identity	% Similarity	Clone ID
1	574130	(AJ223500) nidogen-2 [Homo sapiens] Length = 1375	gnl P D c 1237850	3 716	87	87	11013C56
2	637706			3 1025			HJAA154
3	638162			109 696			INTMW23
4	684310			10 300			11FXJA96
5	731016	protease [Human endogenous retrovirus K] >sp P87892 P87892 PROTEASE (FRAGMENT). Length = 334	gnl P D e290663	2 370	66	83	11PLBP54
6	827771			188 322			11P1CR50
7	828193	MAGE-3b [Homo sapiens] >gi 533523 MAGE-6 antigen [Homo sapiens] >gnl P D d1007417 MAGE-6 protein [Homo sapiens]	gi 499122	237 716	97	97	HMMBI07
8	828194			243 401			HPKAA18
9	828199			2 463			HPJCU04
10	828221	put. LAR preprotein (AA -16 to 1881) [Homo sapiens] >pir S03841 TDHULK leukocyte antigen-related protein precursor - human Length = 1897	gi 34267	1 1326	100	100	HWHQP39
11	828235			3 248			11WB3B77
12	828236	Gu protein [Homo sapiens] >pir P6010 P6010 RNA helicase Gu - human (fragment) >sp Q13436 Q13436 NUCLEOLAR RNA HELICASE GU (FRAGMENT). Length = 801	gi 1230504	1 1425	84	84	11WB3P29
13	828237			3 779			11W1PW78

14	828239	(AC002451) pyruvate dehydrogenase kinase isoform 4 [Homo sapiens] >gi 139197 pyruvate dehydrogenase kinase isoform 4 [Homo sapiens]	gi 2337883	2	433	87	87	HWACS81
15	828242	(AF044321) cytochrome c oxidase assembly protein COX11 [Homo sapiens] >gi 3170264 (AF044321) cytochrome c oxidase assembly protein COX11 [Homo sapiens]	gi 3170264	3	731	100	100	HWBAS37
16	828247	(AF109906) NG22 [Mus musculus] Length = 707	gi 3986770	3	554	39	61	HWBBX45
17	828248	M1 subunit of ribonucleotide reductase [Homo sapiens] >gi 36153 large subunit ribonucleotide reductase [Homo sapiens] >pir S16680 S16680 ribonucleoside-diphosphate reductase (EC 1.17.4.1) chain M1 - human Length = 792	gi 36065	254	625	82	82	HWBAJ23
18	828250	put. ribosomal protein L3 (AA 1 - 348) [Homo sapiens] >pir A27294 R5HUL3 ribosomal protein L3 precursor, mitochondrial - human Length = 348	gi 34754	58	408	94	94	HWBBN56
19	828256			393	1193			HWSGZ25
20	828267			3	497			HWIK57
21	828269			214	492			HWBF75
22	828272			89	607			HWYB27
23	828273	(AF047020) alpha-methylacyl-CoA racemase [Homo sapiens] >sp O43673 O43673 ALPHA-METHYLACYL-COA RACEMASE (EC 5.1.99.4). Length = 380	gi 2896148	300	539	79	89	HWLCJ25
24	828290			648	914			HWGH59
25	828326	Ki antigen [Mus musculus] >gnl PID d1029778 (AB007139) PA28 gamma subunit [Mus musculus] >sp O35563 O35563 KI ANTIGEN. Length = 254	gnl PID d1022900	2	970	99	99	HTXJJ72

26	828397	smooth muscle myosin light chain kinase, smMLCK (C-terminal) [sheep, myometrial tissue, day 127 of gestation, Peptide Partial, 438 aa] [Ovis aries] Length = 438	hbsl175341	1	942			HL.YCG48
27	828405			37	579	98	100	HL.DBK03
28	828461	fra-1 gene product (AA 1-271) [Homo sapiens] >pirS15750 S15750 transforming protein (fra-1) - human >sp P15407 FRA1_HUMAN FOS-RELATED ANTIGEN 1. Length = 271	gil31463	1	873	71	71	HSKEI92
29	828482	Gephyrin [Rattus norvegicus] >pir JH0681 JH0681 gephyrin - rat >sp Q03555 GEPH_RAT GEPHYRIN (PUTATIVE GLYCINE RECEPTOR-TUBULIN LINKER PROTEIN). Length = 736	gil56312	2	940	98	98	HSIGE72
30	828488	BS4 peptide [Mus musculus] >sp P54729 BS4_MOUSE BS4 PROTEIN. Length = 677	gil863014	64	189			HSDJR78
31	828491			386	586			HSDFC18
32	828492			51	212			HSIDGQ64
33	828494			428	733			HSDIC05
34	828496			3	1097	85	93	HSBAV13
35	828498	14.5 kDa translational inhibitor protein, p14.5 [Homo sapiens] Length = 137	gnl PID e1240168	63	500	100	100	HSDXA60
36	828504	CCAAT-box-binding factor [Homo sapiens] >pir A36368 A36368 transcription factor CBF, CCAAT-binding - human	gil179969	173	412			HSAAQ28
37	828507			286	462			HSBCA90
38	828512			3	611	82	82	HSAAV04
39	828516	histone H2A [Homo sapiens] >gil 2062704 histone 2A-like protein [Homo sapiens] >gil 2088554 histone 2A-like protein [Homo sapiens]	gnl PID e268230	36	458	100	100	HSBAL82

40	828519	DEAD box-like RNA helicase [Arabidopsis thaliana] >sp O23251 O23251 DEAD BOX-LIKE RNA HELICASE (FRAGMENT). Length = 450	gnl PID e1316345	142	474	38	58	IIRGIBQ34
41	828521			31	531			IIRGDE67
42	828522	Unknown	gi 6332974	361	684	99		HIROBP89
43	828525	cytokine receptor [Homo sapiens] >sp Q14213 Q14213 CYTOKINE RECEPTOR PRECURSOR.		14	463		99	HIRGTJ13
44	828529	ORF_f506 [Escherichia coli] >gi 1789453 (AE000389) aerotaxis sensor receptor, flavoprotein [Escherichia coli]	gi 882594	379	852			HIROEB35
45	828530			134	253			HIRACZ50
46	828536			84	272			IIPYSC02
47	828537			1	270			HPZAA72
48	828539			130	279			HPWDG48
49	828540			3	278	100	100	IIPWCG66
50	828542	(AF093263) homer-2a [Homo sapiens] >sp G383461 G3834617 HOMER-2A. Length = 343	gi 3834617	366	626	96	97	HRAAA23
51	828543			3	554			HPWCSI4
52	828544			277	474			IIPWDE02
53	828546			1	1302			IIPWBZ53
54	828550			61	147			IIPWBR41

55	828551	prostate-specific membrane antigen [Homo sapiens] >pir A5688 A5688 prostate-specific membrane antigen - human	gi 190664	61	585			11PWCG88
56	828553	NF-IL6-beta protein [Homo sapiens] >pir A40225 A40225 transcription activator NF-IL6 beta - human Length = 269		2	655	95	95	11PWCG57
57	828557	T-cell receptor (V-J-C) precursor [Homo sapiens] >pir A26659 A26659 T-cell receptor gamma-1 chain C region - human {SUB 138-310}	gi 189176	3	359	100	100	11PTVR29
58	828560	>gi 339080 T cell receptor gamma chain [Homo sapiens] {SUB 139-310} >gi 339089 T-cell receptor gamma-chain constant region [Homo sapiens]	gi 339400	381	683	100	100	11PWAY42
59	828561	zinc finger protein [Homo sapiens] >pir S4707 S47071 finger protein 11ZF3, Krueppel-related - human (fragment)	gi 498725	1	204	96	96	11PW13S62
60	828565	thyroid receptor interactor [Homo sapiens] Length = 286		3	962			11PWAZ16
61	828566	envelope protein [Woodchuck hepatitis B virus] >pir A03708 SAVLC2 large surface antigen - woodchuck hepatitis virus (clone 2) Length = 431		1214	1423			11PW AJ41
62	828567			204	440			11PRT P24
63	828568		gi 703112	2	475	97	100	11PRSB55
64	828569		gi 336133	204	395	38	47	11PWBR81
65	828570	DY3.6 [Caenorhabditis elegans]		380	580			11PRT140
66	828571	>sp O45323 O45323 DY3.6 PROTEIN. Length = 379	gn P1D c1345081	2	670	27	61	11PRT180
67	828574	rTSbeta [Homo sapiens] >sp Q15407 Q15407 RTSBETA. Length = 416	gn P1D c189422	3	458	89	89	11PRTS71
68	828575			3	209			11PRTI65

69	828577	phospholipase A2 [unidentified] >gi 190887	gi 833246	135	395	89	89	HPRTQ68
70	828578	synovial phospholipase A-2 [Homo sapiens] >gi 190889 synovial phospholipase A-2 (EC 3.1.1.4) [Homo sapiens] >pir A32862 PSHUYF phospholipase A2 (EC 3.1.1.4) precursor, synovial fluid - human >sp P14555 PA2M_HUMAN		136	627			HPRTCL59
71	828580			2	340			IPRCS86
72	828581			103	339			IPRSD02
73	828583			258	419			IPRTL26
74	828585	HOXB13 [Homo sapiens] Length = 284 (AF043431) retinoblastoma-interacting protein [Homo sapiens] >sp O75371 O75371	gi 1764090	1	285	100	100	IPRCN60
75	828587	RETINOBLASTOMA-INTERACTING PROTEIN. Length = 897	gi 3452281	139	534	100	100	IPRCT61
76	828590			120	248			IPRCL51
77	828592	breakpoint cluster region protein [Homo sapiens] >sp Q12844 Q12844 BREAKPOINT CLUSTER REGION PROTEIN (FRAGMENT). Length = 889	gi 487346	48	611	98	98	IPRCT63
78	828593	XP-G factor [Homo sapiens] >pir S35993 S35993 DNA repair protein XPGC - human >sp G303059 G303059 XPGC=DNA REPAIR PROTEIN RAD2 HOMOLOG. (SUB 1166- 1186) Length = 118	gi 298111	1	1272	87	87	HPRTJ39
79	828594			84	353			HPRCM59
80	828596	homeobox protein [Homo sapiens] >pir S19010 S19010 homeotic protein PBX3a - human >sp P40426 PBX3_HUMAN PRE-B- CELL LEUKEMIA TRANSCRIPTION FACTOR-3 (HOMEBOX PROTEIN PBX3).	gi 295315	1	213	93	93	IPRCH15

Length = 434

81	828597	(AL031532) yeast gtr2 homolog, novel small GTPase subfamily protein [Schizosaccharomyces pombe] >sp O74544 O74544 YEAST GTR2 HOMOLOG, NOVEL SMALL GTPASE SUBFAMILY PROTEIN. Length = 31	gnl PIDe1319429	1	903	70	85	IIPRBB367
82	828598	acid phosphatase [Homo sapiens] Length = 386 prostate-specific membrane antigen [Homo sapiens] >pir A56881 A56881 prostate-specific membrane antigen - human seminal plasma protein precursor [Homo sapiens] >gi 514372 beta-microseminoprotein [Homo sapiens] >gi 825707 prostatic secretory protein (PSP-94) [Homo sapiens] prostatic acid phosphatase [Homo sapiens] >gi 189621 acid phosphatase [Homo sapiens] >gi 515997 prostatic acid phosphatase [Homo sapiens]	gi 189619 gi 190664	1	108			IIPRAX93
83	828601			2	520			IIPRT175
84	828605			383	601			IIPRAY38
85	828608			21	533	95	96	IIPRBF14
86	828609			186	899	100	100	IIPRBH58
87	828610		gi 338415	3	398	100	100	IIPRTJ08
88	828617		gi 189613	3	350			IIPRAD26
89	828620			3	650	94	94	IIPRBF16
90	828621		gi 189613	4	126			IIPRAG37
91	828622			28	156			IIPRAQ51
92	828623			125	313			IIPRAG59
93	828625			87	275			IIPRA122
94	828632			68	406			IIPQBV63
95	828635			916	1344			IIPMGE79

96	828637	(AC005600) PKD1 [Homo sapiens] >sp O75276 O75276 PKD1 (FRAGMENT). Length = 1339	gi 3522923	1	366	70	71	HP0A153
97	828639			72	158			HPMDDB85
98	828645			2	313			HPJCK50
99	828648	(AF059569) actin binding protein MAYVEN [Homo sapiens] >sp G3789797 G3789797 ACTIN BINDING PROTEIN MAYVEN. Length = 593	gi 3789797	210	677	32	48	HPJBV55
100	828649	neuropeptide Y [Homo sapiens] >gi 189282 neuropeptide Y [Homo sapiens] >gi 2992498 (AC004485) neuropeptide Y precursor [Homo sapiens]	gi 189274	121	375	100	100	HPWBU56
101	828651	similar to ATPases associated with various cellular activities (AAA);	gi P1D1e1351769	41	742	51	69	HPJDA05
102	828652			1	189			HPJCY65
103	828655			60	251			HPJBW32
104	828657	(AF061283) neuronal protein 4.1 [Mus musculus] >sp G3790545 G3790545 NEURONAL PROTEIN 4.1. Length = 879	gi 3790545	38	328	45	67	HPJBD30
105	828660			103	231			HPJCL80
106	828663	calnexin [Homo sapiens] >gi 186523 calnexin [Homo sapiens] >pir A46673 A46673 calnexin precursor - human >sp P27824 CAI.X_HUMAN CALNEXIN PRECURSOR (MAJOR HISTOCOMPATIBILITY COMPLEX CLASS I ANTIGEN-BINDING PROTEIN P88) (P90) (IP90). Length = 592	gi 306481	41	703	87	87	HPJCT42
107	828666			1	246			HPJBI71
108	828668			61	315			HPJBK31
109	828669			1	225			HPJBU60
110	828670			222	350			HPICC36

21

111	828671	(AJ005866) Sqv-7-like protein [Homo sapiens] >sp E1360006 E1360006 SQV-7-LIKE PROTEIN (FRAGMENT). Length = 261	gnl PID e1360006	3	1025	89	90	IIPJAD23
112	828672			1	255			HPICD86
113	828675	MCM4 [Homo sapiens] >sp Q2754697 Q2754697 MCM4 (FRAGMENT). Length = 712	gi 2754697	2	2173	99	99	HPJBZ66
114	828677			113	268			HPICC05
115	828678	SNAP43 [Homo sapiens] >gi 1174203 PSE- binding factor PTF gamma subunit [Homo sapiens] >pir JC608 JC608 proximal sequence element-binding transcription factor gamma chain - human >sp Q16533 Q16533 PSE-BINDING FACTOR PTF GAMMA SUBUNIT. Length = 368	gi 623244	2	664	98	98	HPJAA76
116	828679			142	318			IIPJAC93
117	828680	DNA primase (subunit p48) [Homo sapiens] >pir S45630 S45630 DNA primase chain p48 - human >sp P49642 P49642 HUMAN DNA PRIMASE SMALL SUBUNIT (EC 2.7.7.-) (DNA PRIMASE 49 KD SUBUNIT) (P49). >gi 2353692 DNA primase 1 [Homo sapiens] {SUB 97-146} Length = 420	gi 510406	74	652	100	100	HPICG94
118	828681			3	167			HPJAA30
119	828682			3	617			HPIBM51
120	828683			54	329			HPIBR22
121	828686	(AF006010) progesterin induced protein [Homo sapiens] >sp G4101695 G4101695 PROGESTIN INDUCED PROTEIN. Length = 2796	gi 4101695	2	886	95	97	HPIBQ56
122	828687			27	131			HPIBS12

123	828688	CCAAAT-box DNA binding protein subunit NF-YB [Homo sapiens] >sp P25208 CBFA_HUMAN CCAAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B). creatine kinase [Homo sapiens] >pir A31431 A30789 creatine kinase (EC 2.7.3.2) precursor, mitochondrial - human >sp P12532 KCRU_HUMAN CREATINE KINASE, UBIQUITOUS MITOCHONDRIAL PRECURSOR (EC 2.7.3.2) (U-MTCK) (MIA-CK) (ACIDIC-TYPE MITOCHONDRIAL CREATINE K (AJ223301) aralkyl acyl-CoA:amino acid N-acyltransferase [Bos taurus] >gi 2865607 (AI:045032) aralkyl acyl-CoA:amino acid N-acyltransferase [Bos taurus] >sp O46686 O46686 ARALKYL ACYL-COA:AMINO ACID N-ACYLTRANSFERASE (EC 2.3.1.13) (GLYCINE N-ACYLTRANSFERASE dJ1409.2 (Melanoma-Associated Antigen MAGE LIKE) [Homo sapiens] >sp O76058 O76058 DJ1409.2 (MELANOMA-ASSOCIATED ANTIGEN MAGE LIKE). Length = 606	128	757	100	100	100	11PJA20
124	828689	creatine kinase [Homo sapiens] >gi 180590	227	1222	84	84	84	HPICC13
125	828692	aralkyl acyl-CoA:amino acid N-acyltransferase [Bos taurus] >gi 2865607 (AI:045032) aralkyl acyl-CoA:amino acid N-acyltransferase [Bos taurus] >sp O46686 O46686 ARALKYL ACYL-COA:AMINO ACID N-ACYLTRANSFERASE (EC 2.3.1.13) (GLYCINE N-ACYLTRANSFERASE dJ1409.2 (Melanoma-Associated Antigen MAGE LIKE) [Homo sapiens] >sp O76058 O76058 DJ1409.2 (MELANOMA-ASSOCIATED ANTIGEN MAGE LIKE). Length = 606	278	1000	49	70	70	HPIBO30
126	828693	aralkyl acyl-CoA:amino acid N-acyltransferase [Bos taurus] >sp O46686 O46686 ARALKYL ACYL-COA:AMINO ACID N-ACYLTRANSFERASE (EC 2.3.1.13) (GLYCINE N-ACYLTRANSFERASE dJ1409.2 (Melanoma-Associated Antigen MAGE LIKE) [Homo sapiens] >sp O76058 O76058 DJ1409.2 (MELANOMA-ASSOCIATED ANTIGEN MAGE LIKE). Length = 606	1	426	45	69	69	HPIBL27
127	828694	aralkyl acyl-CoA:amino acid N-acyltransferase [Bos taurus] >gi 2865607 (AI:045032) aralkyl acyl-CoA:amino acid N-acyltransferase [Bos taurus] >sp O46686 O46686 ARALKYL ACYL-COA:AMINO ACID N-ACYLTRANSFERASE (EC 2.3.1.13) (GLYCINE N-ACYLTRANSFERASE dJ1409.2 (Melanoma-Associated Antigen MAGE LIKE) [Homo sapiens] >sp O76058 O76058 DJ1409.2 (MELANOMA-ASSOCIATED ANTIGEN MAGE LIKE). Length = 606	1	333				HPIBY69
128	828696	aralkyl acyl-CoA:amino acid N-acyltransferase [Bos taurus] >gi 2865607 (AI:045032) aralkyl acyl-CoA:amino acid N-acyltransferase [Bos taurus] >sp O46686 O46686 ARALKYL ACYL-COA:AMINO ACID N-ACYLTRANSFERASE (EC 2.3.1.13) (GLYCINE N-ACYLTRANSFERASE dJ1409.2 (Melanoma-Associated Antigen MAGE LIKE) [Homo sapiens] >sp O76058 O76058 DJ1409.2 (MELANOMA-ASSOCIATED ANTIGEN MAGE LIKE). Length = 606	171	347				HPIBA33

129	828697	kynurenine/alpha-aminoadipate aminotransferase [Rattus norvegicus] >sp Q64602 Q64602 KYNURENINE/ALPHA-AMINOADIPATE AMINOTRANSFERASE (EC 2.6.1.7) (KYNURENINE--OXOGLUTARATE AMINOTRANSFERASE) (KYNURENINE AMINOTRANSFERASE). Length = 425	gij1050752	258	422	61	72	IIPIC1303
130	828699			3	1109			IIPBI.48
131	828702	prostate- specific membrane antigen [Homo sapiens] >pir A56881 A56881 prostate-specific membrane antigen - human >bbs 16419 prostate-specific membrane antigen,	gij190664	118	744	76	78	IPIAZ02
132	828703			285	689			IPIBI396
133	828704	put. DNA topoisomerase I (AA 1-864) [Escherichia coli] >gn PID d1015527 DNA topoisomerase I (EC 5.99.1.2) (w-protein) (Relaxing enzyme) (Untwisting enzyme) (Swivelase). [Escherichia coli] mitotic centromere-associated kinesin [Homo sapiens] >sp Q99661 Q99661 MITOTIC CENTROMERE-ASSOCIATED KINESIN. Length = 725	gij415338	2	406	98	98	IPIBI330
134	828706		gij1695882	559	1788	98	98	IPIBI311
135	828708			2	589			IPIAW81
136	828711			1	93			IPIAZ32
137	828712			49	309			HPIAU16
138	828713			142	396			IPIAV37
139	828714			68	1849			IPIAV20
140	828715			174	356			IPIAS34
141	828718	ipa-6d gene product [Bacillus subtilis] >gn PID e1186348 alternate gene name: ipa-6d; similar to quinone biosynthesis [Bacillus subtilis]	gij413930	403	1308	35	57	IPIAL41

142	828723	UDP glucuronosyltransferase precursor [Homo sapiens] >pir A48633 A48633 dihydrotestosterone/androstenediol UDP-glucuronosyltransferase isoform 3, udpgth-3 - human	gj 475759	3	206	97	100	HIPIAL34
143	828726	hydrophobic membrane-bound protein [Escherichia coli] >gj 147818 part of a molybdenum periplasmic binding protein dependent transport system [Escherichia coli] >gj 973215 ModB [Escherichia coli] (AF044954) NADH:ubiquinone oxidoreductase PDSW subunit [Homo sapiens] >gj 4165091 (AF088991) NADH-ubiquinone oxidoreductase PDSW subunit [Homo sapiens] Length = 172 MAK11 protein [Saccharomyces cerevisiae] >gj 486013 ORF YKL021c [Saccharomyces cerevisiae] >pir A29938 A29938 MAK11 protein - yeast (Saccharomyces cerevisiae) >sp P20484 MK11_YEAST MAK11 PROTEIN. Length = 468	gj 504499	1	255	98	98	HIPIAS69
144	828728		gj 4164442	1	498	84	86	HIPIAS40
145	828730		gj 171877	394	1569	34	64	HIPIAF82
146	828732	rab geranylgeranyl transferase [Homo sapiens] >pir JC5538 JC5538 Rab geranylgeranyl transferase (EC 2.5.1.-) alpha chain - human >sp E1256376 E1256376 RAB GERANYLGERANYL TRANSFERASE. Length = 567	gnl PID e1256376	155	868	97	97	HIPIAN07
147	828733	(AF006265) cancer associated surface antigen [Homo sapiens] >gnl PID d1023440 (AB007619) EBAG9 [Homo sapiens] >sp O00559 O00559 CANCER ASSOCIATED SURFACE ANTIGEN. Length = 213	gj 2213934	202	438	90	90	HIPIAK81
148	828735			369	1139			HIPIAE30

149	828736	glandular kallikrein precursor [Homo sapiens] >pir A29586 A29586 tissue kallikrein (EC 3.4.21.35) HGK-1 precursor - human >sp P20151 KLK2_HUMAN GLANDULAR KALLIKREIN 2 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN) (PROSTATE) (HGK-1). Length = 261 serine/threonine kinase [Rattus norvegicus] >sp O08678 O08678 SERINE/THREONINE KINASE. Length = 793 androgen regulated homeobox protein [Homo sapiens] >sp Q99801 HK31_HUMAN HOMEBOX PROTEIN NKX-3.1. Length = 234	gi 386842	1	132	HPFEA11
150	828739			60	347	HPFAA46
151	828740			2	394	HPFAC69
152	828742			2	475	HPHAB61
153	828748			3	707	HPFAB20
154	828749	serine/threonine kinase [Rattus norvegicus] >sp O08678 O08678 SERINE/THREONINE KINASE. Length = 793 androgen regulated homeobox protein [Homo sapiens] >sp Q99801 HK31_HUMAN HOMEBOX PROTEIN NKX-3.1. Length = 234	gn IPID e290956	443	826	HPFAA79
155	828752		gi 1732378	1051	1692	HPFAA91
156	828753	cytochrome c oxidase subunit VIc preprotein [Homo sapiens] >gi 3859868 (AF067637) cytochrome c oxidase subunit VIc [Homo sapiens]	gn IPID e223120	2	187	HPFEA08
157	828754			423	566	HPFDD83
158	828757			2	409	HPFDI21
159	828761			3	113	HPFDE61
160	828762			3	317	HPFDE33
161	828764			51	329	HPMSI48
162	828765	cytochrome c oxidase subunit VIc preprotein [Homo sapiens] >gi 3859868 (AF067637) cytochrome c oxidase subunit VIc [Homo sapiens]	gn IPID e223120	3	80	HPFDB49
163	828766			90	242	HPFDT61
164	828767			797	937	HPWDK71
165	828768			1109	1324	HPFDD04
166	828770			156	392	HPFDF79

167	828771	(AF001629) WASP interactor protein [Homo sapiens] >sp G4100621 G4100621 WASP INTERACTOR PROTEIN (FRAGMENT). Length = 328	gi 4100621	1	273	55	61	HPFD550
168	828772			200	340			HPFD128
169	828773			115	348			HPFD85
170	828775			23	208			HPFCR19
171	828776			3	134			HPFCY40
172	828777			131	919			HPFDM39
173	828778			2	121			HPFCZ89
174	828780			46	420			HPFDA70
175	828781			408	734			HPFCP06
176	828782			61	186			HPFDI40
177	828783	relaxin [Homo sapiens] >gi 490063 H1-relaxin [Homo sapiens] >gi 412167 relaxin [Homo sapiens] >gi 512431 preprorelaxin [Homo sapiens] >gi 35933 prepro-relaxin H1 [Homo sapiens]	gi 490056	68	253	70	70	HPFCH80
178	828784			82	321			HPFC179
179	828785			32	250			HPFCX77
180	828786			302	532			HPFC131
181	828788			341	538			HPFC159
182	828790			195	317			HPFC153
183	828791			6	140			HPFC114
184	828792			121	801			HPFCC91
185	828794			1219	1440			HPFCJ56
186	828797			128	259			HPFCC42
187	828798			237	350			HPFC176
188	828799			113	322			HPFAA95
189	828801			90	239			HPFAG41
190	828802			165	392			HPFC126

191	828803	(AB022017) AMP-activated protein kinase alpha-1 [Homo sapiens] >sp D1037533 D1037533 AMP-ACTIVATED PROTEIN KINASE ALPHA-1. >gnl PID e315274 AMP-activated protein kinase alpha-1 [Homo sapiens] {SUB 294-550}	gnl PID d1037533	96	458	83	83	IIP1BA83
192	828804			98	286			IIP1AC32
193	828805			166	303			IIP1CF17
194	828807			1	195			IIP1CF96
195	828809			147	236			IIP1AC52
196	828810			1	153			IIP1BT31
197	828811			283	426			IIP1AA06
198	828817			2	160			IIP1CAC47
199	828818			1	258			IIP1AA76
200	828819			345	623			IIP1BG44
201	828820			314	502			IIP1AB80
202	828821			246	416			IIP1CAF64
203	828823	spore coat protein SP87 [Dictyostelium discoideum] Length = 677	gi 915203	267	875	44	61	IIP1AB79
204	828824			458	643			IIP1AC56
205	828825			132	446			IIP1DDY72
206	828826			2	730			IIP1CAN60
207	828829			499	672			IIP1CA054
208	828830	Arnt [Homo sapiens] >pir 59550 59550 Arnt - human >sp P27540 ARNT_HUMAN ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR (ARNT PROTEIN) (DIOXIN RECEPTOR, NUCLEAR TRANSLOCATOR) (HYPOXIA-INDUCIBLE FACTOR 1 BETA) (HIF-1 BETA). Length = 789	gi 179004	1	219	90	92	IIP1AA27
209	828833			42	278			IIP1CAB16
210	828835			61	474			IIP1DC43

211	828838	chordin [Xenopus laevis] >pir A55195 A55195 chordin precursor - African clawed frog >sp Q91713 CHRD_XENLA_CHORDIN PRECURSOR (ORGANIZER-SPECIFIC SECRETED DORSALIZING FACTOR). Length = 941	gi 603945	2	1468	43	56	IIPCA032
212	828840			536	679			HOVCJ65
213	828845			69	212			HOSDG69
214	828846			3	1034			HSPBQ12
215	828847			36	395			IIPEAA46
216	828849	(AF041474) BAF53a [Homo sapiens] >sp G4001803 G4001803 BAF53A. Length = 429	gi 4001803	62	1468	100	100	HOVCJ86
217	828850	putative [Homo sapiens] >pir A49364 A49364 59 protein, brain - human (fragment) >sp Q09019 DMR9_HUMAN DMR-N9 PROTEIN (PROTEIN 59) (FRAGMENT). Length = 553	gi 306712	2	283	97	97	IIQICP33
218	828852			96	437			HOAZ63
219	828853	(AC004449) R33683_3 [Homo sapiens] >sp O60372 O60372 R33683_3 (FRAGMENT). Length = 103	gi 2979531	1	465	40	62	HOSAV36
220	828857	uridine kinase [Mus musculus] Length = 260						
221	828861	enhancer of filamentation 1 [Homo sapiens] >gi 490787 Crk-associated substrate related protein Cas-L [Homo sapiens] >sp Q14511 Q14511 ENHANCER OF FILAMENTATION 1. Length = 834	gi 471981	3	1013	74	88	HOQBM19
222	828866	pericentriol material 1 [Homo sapiens] >pir A54103 A54103 centrosome autoantigen PCM-1 - human >sp Q15154 Q15154	gi 280212	2	991	100	100	HPEAES5
223	828872		gi 450277	143	637			IOHBF14
				295	879	93	94	HOHAL47

PERICENTRIOL MATERIAL 1. Length = 2024

224	828874	histone H1(0) (aa 1-194) [Homo sapiens] >pir A24850 HSHU10 histone H1-0 - human >sp P07305 H10_HUMAN HISTONE H1' (H1.0) (H1(0)). (SUB 2-194) Length = 194	gi 32107	3	902	82	82	HOGBL72
225	828875	myosin VI [Homo sapiens] >sp G2304981 G2304981 MYOSIN VI. Length = 1262	gi 2304981	1	450	99	99	HOGCC24
226	828877	75 kDa subunit NADH dehydrogenase precursor [Homo sapiens] >pir S17854 S17854 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 75K chain precursor - human	gi 38079	24	275	95	97	HOFMJ67
227	828878	S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50) old gene name 'AMD' [Homo sapiens] >pir A31786 DCHUDM adenosylmethionine decarboxylase (EC 4.1.1.50) precursor - human	gi 178518	282	1325	95	95	HOGRC089
228	828879	product possesses binding site dependent transcriptional suppressing activity [Homo sapiens] >pir A44351 A44351 transcription repressor E4BP4 - human >sp Q14211 Q14211 E4BP4 GENE. Length = 462	gi 30956	2 139 173	271 969 1639	94	95	HOEJ117
229	828881							HOGAI'39
230	828885							HOEECS8
231	828886	ZNF127-Xp [Homo sapiens] >sp Q13434 Q13434 ZNF127-XP. Length = 485	gi 304599	82 2	228 1327	56	76	HODGT65
232	828887							HOEEN41

233	828889	neurofibromin [Homo sapiens] >sp P21359 NF1_HUMAN NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED) PROTEIN NF-1). >gi 736765 neurofibromatosis 1 [Homo sapiens] {SUB 751-1611} >gi 189161 neurofibromatosis protein type 1 [Homo sapiens] {SUB 1168-1566}	gi 292354	265	690	89	89	110DAQ30
234	828891	FAST kinase [Homo sapiens] >pir I37386 I37386 FAST kinase - human >sp Q14296 Q14296 FAST KINASE. Length = 549	gi 1006659	84	1238	100	100	HODDG78
235	828899	MAP KINASE-ACTIVATED PROTEIN KINASE 2 (EC 2.7.1.-) (MAPK-ACTIVATED PROTEIN KINASE 2) (MAPKAP KINASE 2) (MAPKAPK-2). Length = 400 zinc finger protein 7 (ZFP7) [Homo sapiens] >pir A34612 A34612 zinc finger protein ZNF7 - human Length = 686 RNA helicase [Homo sapiens] >pir S71758 S71758 DEAD box protein MrDb, Myc-regulated - human >sp Q92732 Q92732 RNA HELICASE. Length = 610	splP49137 MKK2_HUMA N	3	344			IINWAA42
236	828907			3	566			HINTSS75
237	828911			1217	1501			HNTMC68
238	828914			586	1176	98	99	HINTRL23
239	828917	zinc finger protein 7 (ZFP7) [Homo sapiens] >pir A34612 A34612 zinc finger protein ZNF7 - human Length = 686 RNA helicase [Homo sapiens] >pir S71758 S71758 DEAD box protein MrDb, Myc-regulated - human >sp Q92732 Q92732 RNA HELICASE. Length = 610	gi 340446	790	1536	57	70	IINTCK38
240	828921		gn PID e254454	123	1253	90	90	IINTRO07
241	828922	(AE000180) biotin synthesis, sulfur insertion? [Escherichia coli] >gi 490219 BIOB gene product [Escherichia coli] >gn PID e305036 BIOTIN SYNTHASE [Escherichia coli] >pir JC2517 SYECBB biotin synthetase (EC 2.8.1.-) - Escherichia coli	gi 1786992	138	1403	95		HINTAB76
242	828924			1	78	95	95	HNHAG14

30

>sp|P12996|BIOB_ECOL

243	828925	cascin kinase I-alpha [Homo sapiens] >pir A57011 A57011 casein kinase I-alpha - human Length = 337 (AL021366) cICK0721Q.3 (Kinesin related protein) [Homo sapiens] >sp O60887 O60887 CICK0721Q.3 (KINESIN RELATED PROTEIN). >gnl PIDe1332987 (AJ010479) kinesin-like protein [Homo sapiens] {SUB I- 274} Length = 673 apurinic/apurimidinic endonuclease [Homo sapiens] >gil183780 apurinic/apurimidinic endonuclease [Homo sapiens] >gil32022 AP endonuclease I [Homo sapiens] >bbs111437 Ref-1=redox factor [human, Peptide, 318 aa] [Homo sapiens] >pir S23550 S23550 DNA- (apurin	376	426			IINGKM39
244	828926		28	522			HNTBH70
245	828928		1	330			HNGNK23
246	828930		412	1467	89	91	HINFJ194
247	828935	pol polyprotein - Moloney murine leukemia virus (strain 3-1R) (fragment) Length = 559	2	1447	86	86	HINTRL26
							31
248	828937	pol polyprotein - Moloney murine leukemia virus (strain 3-1R) (fragment) Length = 559	124	1158	95	95	IINTNM15
249	828940	pol polyprotein - Moloney murine leukemia virus (strain 3-1R) (fragment) Length = 559	1399	1806	58	71	IINGGG72
250	828942		3	386			IINFHK65

1/23/2006, EAST Version: 2.0.1.4

261	828984	high mobility group box [Homo sapiens] >pir A41976 A41976 structure-specific recognition protein, SSRP1 - human Length = 709	gij 184242	322	2388	97	97	IIMUAQ01
262	828985	Similarity to Yeast MSP1 protein (TAT-binding homolog 4) (SW:MSP1_YEAST)		734	928			IIMSG125
263	828988	[Caenorhabditis elegans] >sp P54815 MSP1_CAEEL MSP1 PROTEIN HOMOLOG. Length = 357	gnl PID e 1347884	1	1137	79	88	IIMUBL18
264	828993			78	308			IIMTMB67
265	828995			653	1567			IIMSV02
266	829000			296	478			IIMBW26
267	829005			1	531			IIMQAI48
268	829009	GTP-binding protein [Homo sapiens] >sp O43824 O43824 GTP-BINDING PROTEIN. Length = 442	gnl PID e 1227622	64	927	88	88	IIMQAI69
269	829010	(AF035537) DNA polymerase zeta [Homo sapiens] Length = 3052	gij 2665742	282	1262	93	93	IIMSG1189
270	829012	ribophorin II precursor - human Length = 631	pir B26168 B26168	161	2188	94	95	IIMSJH16
271	829013			1339	1506			IIMAX25
272	829019			41	223			IIMIAJ48
273	829020	similar to WD domain, G-beta repeats (2 domains);	gnl PID e 1345001	21	800	60	77	IIMELR71
274	829021			356	640			IIMIAJ26
275	829026	RIZ [Homo sapiens] >sp Q13029 Q13029 ZINC FINGER PROTEIN RIZ. >pir I38902 I38902 retinoblastoma-binding protein RIZ - human {SUB 3-1721} Length = 1721	gij 3645905	89	1183	87	87	HMELM45

33

276	829030	chaperonin-like protein [Homo sapiens] >pir S48087 S48087 t-complex-type molecular chaperone CCT6 - human >gi 184462 chaperonin-like protein [Homo sapiens] {SUB 143-531} Length = 531	gi 517065	1	1674	95	95	IIMICQ08
277	829035	(AF082516) l-l receptor candidate protein [Homo sapiens] >sp G3462807 G3462807 l-l RECEPTOR CANDIDATE PROTEIN. >gi 3493225 (AF058290) imidazoline receptor antisera-selected protein [Homo sapiens] {SUB 469-1063} Length = 1504	gi 3462807	2	679	98	98	IIMEFK17
278	829041	pyrroline-5-carboxylate reductase [Homo sapiens] >pir A41770 A41770 pyrroline-5- carboxylate reductase (EC 1.5.1.2) - human >sp P32322 PROC_HUMAN PYRROLINE-5- CARBOXYLATE REDUCTASE (EC 1.5.1.2) (P5CR) (P5C REDUCTASE). Length = 319	gi 189498	268	1032	99	100	IIMEIQ04
279	829045			2	1771			IIMEKR35
280	829048			115	1467			IIMEJC44
281	829051			2	256			IIMEJB38
282	829052			795	1154			IIMIBD67
283	829057			116	799			IIMEAF61
284	829058			3	536			IIMEER28
285	829059			310	501			IIMDAQ69
286	829061			3	101			HMCFX82
287	829062	(AF095791) TACC2 protein [Homo sapiens] >sp G3777596 G3777596 TACC2 PROTEIN (FRAGMENT). Length = 653	gi 3777596	1417	2622	50	71	IIMCGK90
288	829063	kinesin-like DNA binding protein KID - human Length = 665	pir S62328 S62328	58	1437	83	84	IIMEFI72
289	829064			2	718			IIMADG63

290	829066	37KD protein, similar to Y122-ECOLI [Escherichia coli] >sp Q47535 Q47535 37KD PROTEIN, SIMILAR TO Y122-ECOLI. Length = 424	g l PID d 1013520	600	1427	98	98	1427	IIMAIIX38
291	829068	(AF037204) RING zinc finger protein [Homo sapiens] >g j3387925 (AF070558) RING zinc finger protein RZF [Homo sapiens] >sp O43567 O43567 RING ZINC FINGER PROTEIN. Length = 381	g j2746333	432	1319	84	84	1319	IIMSI192
292	829069	topoisomerase I [Homo sapiens] >g 473581 DNA topoisomerase I [Homo sapiens] {SUB 5-765} >g l PID e 312191 (AL022394) dJ511B24.1 (Topoisomerase I) [Homo sapiens] {SUB 437-765} Length = 765	g j339804	1	207	69	78	207	IILYET39
293	829074				1269				IILYDE91
294	829077				873				IILYFD84
295	829078				907				IILYCP31
296	829079	putative ATP/GTP-binding protein [Homo sapiens] >sp Q92989 Q92989 PUTATIVE ATP/GTP-BINDING PROTEIN. Length = 425	g l 644402	194	382	93	93	382	IILYBT93
297	829085				783				IIMCEJ41
298	829093	26S proteasome-associated pad1 homolog [Homo sapiens] >sp O00487 O00487 26S PROTEASOME-ASSOCIATED PAD1 HOMOLOG. Length = 310	g l 923256	307	1251	100	100	1251	IILYAN96
299	829099	alpha-L-fucosidase precursor (EC 3.2.1.5) [Homo sapiens] >pir A33427 HWHUFA alpha-L-fucosidase (EC 3.2.1.5) 1 precursor, tissue - human >g l PID e 34843 alpha-L-fucosidase [Homo sapiens] {SUB 357-393} Length = 461	g l 78409	2	850	96	96	850	IILTIDK55
300	829101	protein tyrosine phosphatase [Homo sapiens] Length = 415	g l 804750	3	542	100	100	542	IILYAP23

36

301	829102	!!!! ALU SUBFAMILY SQ WARNING ENTRY sp P39194 ALU7_HUMA !!!! Length = 593	3	59	84	94	HLTEO83
302	829103		265	663			III.WAC24
303	829104		316	525			HL.WAX30
304	829109		3	155			HL.TCF21
305	829111		1	333			HL.TGS92
306	829115		2	670			HL.THA72
307	829116		104	265			HL.QDA07
308	829119		144	374			HL.MCG37
309	829120		611	910			HL.TGP61
310	829121		558	698			HL.QCN32
311	829123	aldelyde oxidase [Homo sapiens] >pir A49634 A49634 aldelyde oxidase (EC 1.2.3.1) - human >sp Q06278 ADO_HUMAN ALDEHYDE OXIDASE (EC 1.2.3.1). Length = 1338	7	585	99	99	III.QDA57
312	829126		2	154			III.QCX53
313	829135	beta-D-galactosidase precursor (EC 3.2.1.23) [Homo sapiens] >gil 79423 beta-galactosidase precursor (EC 3.2.1.23) [Homo sapiens] >pir A32688 A32611 beta-galactosidase (EC 3.2.1.23) precursor - human (AJ005458) protein Phosphatase 2C beta [Bos taurus] >sp O62830 O62830 PROTEIN PHOSPHATASE 2C BETA (EC 3.1.3.16). Length = 387	3	2090	98	98	III.QAM57
314	829136		55	1254	95	96	III.TIS28

315	829138	cytochrome b5 [Homo sapiens] >pir A28936 CBHU5 cytochrome b5, mitochondrial form - human >sp P00167 CYB5_HUMAN CYTOCHROME B5. {SUB 2-134} >gil 81229 cytochrome b5 [Homo sapiens] {SUB 87-134} Length = 134	gil 81227	35	499	89	89	HLH.TN31
316	829142	(AF016509) oxidoreductase [Homo sapiens] >sp O14756 O14756 OXIDOREDUCTASE. Length = 317	gil 2338748	2	1135	99	99	HLH.BJ28
317	829148	protein kinase C iota [Homo sapiens] >gil 598225		55	279			HLH.DP51
318	829149	protein kinase C iota [Homo sapiens] >pir A49509 A49509 protein kinase C (EC 2.7.1.-) iota - human	gil 432274	1	783	99	100	HLH.CD11
319	829156	ORF YDL063c [Saccharomyces cerevisiae] >pir S67598 S67598 probable membrane protein YDL063c - yeast (Saccharomyces cerevisiae)	gnl PID e253210	3	347	82	83	HLH.CD19
320	829162	(A17019767) zinc finger protein [Homo sapiens] >sp O75312 O75312 ZINC FINGER PROTEIN. Length = 459	gil 3510462	3	890	88	89	HLH.GDA89
321	829170	complement factor B [Homo sapiens]		2	160			HLH.DBY56
322	829177	>gil 2347133 (AF019413) complement factor B [Homo sapiens] >gil 553536 MHC factor B [Homo sapiens] {SUB 339-509} Length = 764	gil 291922	1	600	86	87	HLH.DBN31
323	829179			518	847			HLH.2AG36

37

324	829184	CDC2 polypeptide (CDC2) (AA 1-297) [Homo sapiens] >gi 29841 CDC2 protein (AA 1-297) [Homo sapiens] >pir A29539 A29539 protein kinase (EC 2.7.1.37) cdc2 - human >sp P06493 CC2_HUMAN CELL DIVISION CONTROL PROTEIN 2 HOMOLOG (EC 2.7.1.-) (P34 PROTEIN KINASE)	gi 29839	553	1005	98	98	111.1B1D94
325	829185	M-phase phosphoprotein 4 [Homo sapiens]		77	295			111.2A1106
326	829188	>sp Q99545 Q99545 M-PHASE PHOSPHOPROTEIN 4 (FRAGMENT). Length = 611	gnl PID e248491	282	1238	92	92	111.AA1B63
327	829190	(AF038869) eukaryotic initiation factor 4E-binding protein 3 [Homo sapiens] >sp O60516 O60516 EUKARYOTIC INITIATION FACTOR 4E-BINDING PROTEIN 3. Length = 100	gi 3169393	3	359	87	87	111.2AG38
328	829193	protein kinase [Homo sapiens] >pir S34130 S34130 serine/threonine-specific protein kinase PLK (EC 2.7.1.-) - human >sp P53350 PLK1_HUMAN SERINE/THREONINE-PROTEIN KINASE PLK (EC 2.7.1.-) (PLK-1) (SERINE- THREONINE PROTEIN KINASE 13) (STPK13). Length = 603	gi 312998	2	988	94	94	111.4AF38
329	829196	TAK1 binding protein [Homo sapiens]		1	432			111.IAR10
330	829197	>sp Q15750 Q15750 TAK1 BINDING PROTEIN. Length = 504	gi 1401126	1	252	75	76	111.IBM07
331	829202	(AF060502) peroxisome assembly protein PEX10 [Homo sapiens] >sp O60683 PEXA_HUMAN PEROXISOME ASSEMBLY PROTEIN PEX10 (PEROXIN-10). Length = 326	gi 3170653	97	465	92	94	111.IAY04

332	829203	cyclin G2 [Homo sapiens] >gi 1236915 cyclin G2	1	258		HL1AL88	39
333	829209	[Homo sapiens] >sp Q16589 Q16589 CYCLIN G2. Length = 344	127	342		HL2AF80	
334	829210		148	315		HL1AG80	
335	829214		2	484	74	HKMSB51	
336	829215						
337	829219		29	175		HL1AG81	
338	829220		24	290		HL1AG22	
339	829222	(AF016371) U-snRNP-associated cyclophilin [Homo sapiens] >gi 3647230 (AF036331) cyclophilin [Homo sapiens] >sp O43447 O43447 U-SNRNP-ASSOCIATED CYCLOPHILIN (EC 5.2.1.8). Length = 177	68	664		HKMMC06	
340	829223		1	549	100	HKGBU67	
341	829225	probable transposase - human transposable element MER37 >pir S72486 S72486 putative transposase - human transposon MER37 (fragment) {SUB 177-349} Length = 454	2	187		HL1AC64	
342	829226	pre-B cell enhancing factor [Homo sapiens] >pir A55927 A55927 pre-B cell enhancing factor - human >sp P43490 PBEF_HUMAN PRE-B CELL ENHANCING FACTOR PRECURSOR. Length = 491	1607	1720	80	HL1B188	
343	829227		186	1730	97	HKMMZ30	
344	829231						
345	829232	cyclin A [Homo sapiens] >gi 510604 cyclin A [Homo sapiens] >pir S08277 S08277 cyclin A - human >sp P20248 CG2A_HUMAN G2/MITOTIC-SPECIFIC CYCLIN A. Length = 432	285	548		HKIYE27	
			42	92		HKMMF67	
			2	1546	94	HKGDC59	
346	829233		123	446		HKGBH49	

347	829239	palmitoyl-protein thioesterase [Homo sapiens] >gi 1314355 palmitoyl protein thioesterase [Homo sapiens] >gi 2465725 (AF022211) palmitoyl-protein thioesterase [Homo sapiens] >sp P50897 PPT_HUMAN PALMITOYL- PROTEIN THIOESTERASE PRECURSOR (EC 3.1.2.22)(PALMI	gi 1160967	141	782	100	100	IKFBA66
348	829240			144	347			IKGAB62
349	829242			2	955			IKI1AK14
350	829246	(AF094583) putative HIV-1 infection related protein [Homo sapiens] >sp G388593 G388593.1 PUTATIVE HIV-1 INFECTION RELATED PROTEIN (FRAGMENT), length = 129	gi 388593.1	68	424	89	89	IKAFK34
351	829250			169	309			IKAJW63
352	829253			158	982			IKAHIA61
353	829256			1043	1831			IKAFIL67
354	829263	histone H4 [Tigriopus californicus] >gi 297562 histone H4 [Chironomus thummi] >gi 7084 histone H4 gene product [Chironomus thummi] >gi 7440 histone H4 [Drosophila hydei] >gn PID e242831 histone H4 [Drosophila hydei] >gn PID e242923 histone H4 [Drosophil similar to S. cerevisiae longevity-assurance protein 1 (SP:P38703)] [Caenorhabditis elegans] >sp Q17870 Q17870 SIMILAR TO S. CEREVISIAE LONGEVITY-ASSURANCE PROTEIN 1. Length = 362	gi 10616	2	361	98	98	HKADJ19
355	829266		gi 123105	115	636	43	58	HKADL80

356	829271	cAMP response element regulatory protein [Homo sapiens] >gnl PID d1014939 TAXREB67 protein [Homo sapiens] >pir A45377 A45377 transcription factor CREB-2 - human >sp P18848 ATF4_HUMAN CYCLIC-AMP- DEPENDENT TRANSCRIPTION FACTOR ATF-4 (DNA-BINDING PROTEIN TAX unknown [Homo sapiens] >pir I38891 I38891 hypothetical protein - human (fragment) >sp Q13021 BENE_HUMAN BENE PROTEIN (FRAGMENT). Length = 148 (AB006202) cytochrome b small subunit of complex II [Homo sapiens] >sp O14521 DHSD_HUMAN SUCCINATE DEHYDROGENASE [UBIQUINONE] CYTOCHROME B SMALL SUBUNIT PRECURSOR (CYBS) (SUCCINATE- UBIQUINONE REDUCTASE MEMBRANE ANCHOR SUBUNIT). Length = 159 Similar to D.melanogaster cadherin-related tumor suppressor [Homo sapiens] >sp Q92566 Q92566 MYELOBLAST KIAA0279 (FRAGMENT). Length = 2408	gi 181041	261	1118	86	86	HLIAG18
357	829273		gi 1000712	1	507	94	94	HKAEP12
358	829274		gnl PID d1022913	55	546	76	76	IIKAP138
359	829276		gnl PID d1014097	272	2422	90	90	IIKAC1358
360	829279	(AC005620) R33590_2, partial CDS [Homo sapiens] >sp O75291 O75291 R33590_2, PARTIAL CDS (FRAGMENT). Length = 121	gi 3548790	163	597	95	95	IIKAA581
361	829280			172	375			IIKSB47
362	829283			235	414			IJAAT37
363	829284			2	322			IIMBB19
364	829285			706	912			HKADQ69
365	829287			134	358			IJAAB39
366	829295			81	212			IJACK32

367	829296	mitotic kinase-like protein-1 [Homo sapiens] >pir S28262 S28262 kinesin-related protein MKLP-1 - human >sp Q02241 MKLP_HUMAN MITOTIC KINESIN-LIKE PROTEIN-1. Length = 960	gij34672	352	666			IISAN67
368	829297			1	225	98	98	IJPBA19
369	829298	O-6-methylguanine-DNA methyltransferase [Homo sapiens] >gij307199 6-O-methylguanine- DNA methyltransferase (EC 2.1.1.63) [Homo sapiens] >gij34559 O-6-methylguanine-DNA methyltransferase [Homo sapiens] >pir A34889 XUHUMC methylated-DNA-- protein-cysteine S-m	gij187579	2	694	88	88	IISAV27
370	829302	putative [Homo sapiens] >pir B41648 B41648 protein-tyrosine-phosphatase (EC 3.1.3.48) cdc25B - human >sp P30305 MPI2_HUMAN M- PHASE INDUCER PHOSPHATASE 2 (EC 3.1.3.48). >gij2739200 (AF036233) cdc25B phosphatase [Homo sapiens] {SUB 56-338} Length = 566	gij180173	600	929			HIBEJ72
371	829304			300	716			HK AAL43
372	829320			161	853	100	100	HIBC185
373	829322	capping protein alpha subunit isoform 1 [Homo sapiens] >pir G02639 G02639 capping protein alpha subunit isoform 1 - human >sp P52907 CAZI_HUMAN F-ACTIN CAPPING PROTEIN ALPHA-1 SUBUNIT (CAPZ). Length = 286	gij1336099	3	938	95	95	IUBCV27
374	829355			3	782			HIIEAA46

42

375	829364	initiation factor 2 alpha [Bos taurus] >gi 204002 translational initiation factor eIF-2, alpha subunit [Rattus norvegicus] >pir A26711 A26711 translation initiation factor eIF-2 alpha chain - rat >pir S18461 S18461 translation initiation factor eIF-2 alph	gi 325	70	651	88	88	I1KAEV74
376	829919	weak similarity to procollagen alpha chain 1(V) chain [Caenorhabditis elegans] >sp Q20220 Q20220 SIMILARITY TO PROCOLLAGEN ALPHA CHAIN 1(V) CHAIN. Length = 697	gi 1065515	272	448	50	74	H1AJAC05
377	829941			215	796			H1AIBC14
378	829945	(AF033188) WSB-2 [Mus musculus] >sp O54929 O54929 WSB-2. Length = 404	gi 2766493	43	222	95	98	H1AGIF36
379	829946			2	319			H1AH CZ18
380	829947			1	1206			H1AICN24
381	829952	HIV-EP2/Schnurri-2 [Homo sapiens] >gi 187405 MHC binding protein-2 [Homo sapiens] {SUB 1184-1323} Length = 1833	gi 182120	478	741	80	82	H1AICL28
382	829954			2	853			H1AGDR03
383	829955	zinc finger protein [Homo sapiens] >sp Q92951 Q92951 ZINC FINGER PROTEIN. Length = 273	gi 1575615	52	885	99	100	H1AGEX65
384	829957	ribosomal protein L22 [Rattus norvegicus] >pir S52084 S52084 ribosomal protein L22 - rat Length = 128	gi 710295	1	744	62	74	H1AGEP17
385	829958			2	418			H1AECI75

43

386	829960	sorbitol dehydrogenase [Homo sapiens] >gi 1755138 sorbitol dehydrogenase [Homo sapiens] >pir A54674 A54674 L-iditol 2-dehydrogenase (EC 1.1.1.14) - human >sp G1755138 G1755138 SORBITOL DEHYDROGENASE. Length = 357 (AF106835) putative DnaJ [Methylovorus sp. strain SS1] >sp G400808 G400808 PUTATIVE DnaJ. Length = 371	gi 520450	2	1069	97	97	HAIBJ62
387	829966	histone H1 [Homo sapiens] >pir S26364 HSHU11 histone H1-1 - human >sp P16403 H1D_HUMAN HISTONE H1D (H1.2). {SUB 2-213} Length = 213	gi 4008081	185	505	40	74	HAGAX57
388	829967	transcription factor ATF-3 - human (fragment) Length = 222	gi 31968	213	542	81	81	HAADD138
389	829970	nuclear RNA helicase [Homo sapiens] >sp O00148 O00148 NUCLEAR RNA HELICASE. Length = 427	pir C34223 C34223	3	878	70	72	HADBH65 HADFU64
390	829981	smooth muscle myosin heavy chain isoform SM1 [human, umbilical cord, fetal aorta, Peptide Partial, 330 aa] [Homo sapiens] >pir 65768 65768 smooth muscle myosin heavy chain isoform SM1 - human (fragment) >sp Q16086 Q16086 SMOOTH MUSCLE MYOSIN HEAVY CHAIN	gi 1905998	26	721	88	88	HAACB061
391	829985		bbs 140615	21	209	100	100	HAACBQ88
392	829986							
393	829988			325	849			HAACAI04
394	829990			266	454			HADFU12

44

395	829991	NGF1-B/nur77 beta-type transcription factor homolog=TINUR [human, T lymphoid cell line, PEER, Peptide, 535 aa] [Homo sapiens] >sp Q1631 Q1631 TINUR=NGF1-B/NUR77 BETA-TYPE TRANSCRIPTION FACTOR HOMOLOG. Length = 535	bbs 164521	2	286	98	98	HACBV53
396	829992	Not56-like protein [Homo sapiens]	gnl PID e276888	289	540			IIACBX74
397	829993	>sp Q92685 NT56_HUMAN NOT56-LIKE PROTEIN. Length = 438		3	440	77	77	II6EDW38
398	829998	(AL033385) dna-directed rna polymerase iii subunit [Schizosaccharomyces pombe]	gnl PID e1339667	270	830	43	65	II6EDK29
399	829999	NNP-1 [Homo sapiens]	gij 2258274	14	142			II6RSE17
400	830000	>sp P56182 NNP1_HUMAN NNP-1 PROTEIN (D21S2056E). Length = 461		545	856	77	77	II6EEQ39
401	830001	homologous to rat HREV107 (ACC.NO. X76453) [Homo sapiens] Length = 162	gij 1054752	397	903	88	88	II2MBY64
402	830005	alpha 1(XVIII) collagen [Mus musculus] >sp Q61437 Q61437 PROCOLLAGEN, TYPE XVIII, ALPHA 1 (ALPHA 1 COLLAGEN) (XVIII) (FRAGMENT). Length = 1288	gij 511298	3	347	37	42	II6EEEX40
403	830009	TFIIIE-beta [Homo sapiens] >bbs 67862 general transcription factor IIE 34 kda subunit, TFIIIE 34 kda subunit [human, Peptide, 291 aa] [Homo sapiens] >pir S29292 S29292 transcription factor TFIIIE-beta - human Length = 291	gij 37070	3	1028	93	93	II2LAD85

404	830010	(AF062346) zinc finger protein 216 splice variant 1 [Homo sapiens] >gi 3643811 (AF062347) zinc finger protein 216 splice variant 2 [Homo sapiens] >gi 3668066 (AF062072) zinc finger protein 216 [Homo sapiens] >sp O76080 O76080 ZINC FINGER PROTEIN 216. >bbs	gi 3643809	1	930	100	100	112MBU62
405	830127	thymopoietin alpha [Homo sapiens] >pir A5574 A5574 thymopoietin alpha precursor - human Length = 694	gi 508725	469	1074	77	78	112MB125
406	830128	subunit of coatomer complex [Homo sapiens] >sp P35606 COPP_HUMAN COATOMER BETA' SUBUNIT (BETA'-COAT PROTEIN) (BETA'-COP) (P102). {SUB 2-906} Length = 906	gi 298097	102 3	770 2234	100	100	112CB1125 112CBU57
407	830129							
408	830137	aldehyde dehydrogenase [Homo sapiens] >sp P30837 DHAS_HUMAN ALDEHYDE DEHYDROGENASE, MITOCHONDRIAL X PRECURSOR (EC 1.2.1.3) (CLASS 2). Length = 517	gi 1263008	2	943	95	95	112CBX43
409	830140	retroviral proteinase-like protein - human (fragment) Length = 165 (AF043735) 14-3-3 epsilon [Bos taurus] >gi 984319 epsilon 14-3-3 protein [Homo sapiens] >gnl PID d1033501 (AB017103) 14-3-3 epsilon [Homo sapiens] >gi 902787 14-3-3 protein epsilon isoform [Homo sapiens] >gi 184725 14-3-3 protein epsilon isoform [Homo sa	pir JE0065 JE0065	347	784	100	100	112CBG30
410	830157		gi 3676399	2	889	99	99	112CBB64

411	830195	90kDa heat shock protein [Homo sapiens] >pirA2946 HHHU84 heat shock protein 90-beta - human >sp P08238 HS9B_HUMAN HEAT SHOCK PROTEIN HSP 90-BETA (HSP 84) (HSP 90). {SUB 2-724} Length = 724	gj 306891	80	631	93	94	HWACG91	
412	830196	90kDa heat shock protein [Homo sapiens] >pirA2946 HHHU84 heat shock protein 90-beta - human >sp P08238 HS9B_HUMAN HEAT SHOCK PROTEIN HSP 90-BETA (HSP 84) (HSP 90). {SUB 2-724} Length = 724	gj 306891	19	1263	100	100	II2CAC90	
413	830409	eIF3-p40 [Homo sapiens] >gj 2351380 translation initiation factor eIF3 p40 subunit [Homo sapiens] >sp O15372 O15372 EIF3-P40. Length = 352	gj 2351380	325	1092	91	91	IIIJCQ28	
414	830417	core protein II precursor [Homo sapiens] >pirA32629 A32629 ubiquinol--cytochrome-c reductase (EC 1.10.2.2) core protein II - human Length = 453	gj 180928	115	744	81	82	IIIMCB154	47
415	830531	5' half of the product is homologues to Bacillus subtilis SAICAR synthetase, 3' half corresponds to the catalytic subunit of AIR carboxylase [Homo sapiens] >pirS14147 S14147 multifunctional purine biosynthesis protein - human Length = 425	gj 28384	112	1059	100	100	IIIMCGQ67	
416	830677	pinin [Canis familiaris] >sp P79149 P79149 PININ. Length = 773	gj 1684845	8	1111	88	88	HLWBS80	
417	831355	GTP-binding protein - mouse Length = 198	pir S39543 S39543	128	730	99	100	IIKMA1333	
418	831420	(AB016869) p70 ribosomal S6 kinase beta [Homo sapiens] >sp D1035383 D1035383 P70 RIBOSOMAL S6 KINASE BETA. Length = 495	gnl PID D1035383	1	672	91	92	IIWBAS06	

419	831702	Gem [Homo sapiens] >pir A54575 A54575 35K GTP-binding protein Gem - human >sp P55040 GEM_HUMAN GTP-BINDING PROTEIN GEM (GTP-BINDING MITOGEN- INDUCED T-CELL PROTEIN) (RAS-LIKE PROTEIN KIR). Length = 296	gil544493	100	1107	93	93	H2LAD84
420	831717	ets2 protein [Homo sapiens] >gi 2736087 (AF017257) erythroblastosis virus oncogene homolog 2 protein [Homo sapiens] >pir B32066 TVHUE2 transcription factor ets-2 - human >sp P15036 ETS2_HUMAN C-ETS-2 PROTEIN. >gi 182271 ets protein [Homo sapiens] {SUB 324	gil182273	278	1309	90	90	HLLBB45
421	832488	tissue-specific secretory protein [unidentified] >gi 32051 HE4 protein [Homo sapiens] >pir S25454 S25454 HE4 protein - human >sp Q14508 EP4_HUMAN MAJOR EPIDIDYMIS-SPECIFIC PROTEIN E4 PRECURSOR (HE4) (EPIDIDYMAL SECRETORY PROTEIN E4). Length = 125 secretory granule proteoglycan peptide core [Homo sapiens] >gi 338062 proteoglycan secretory granule 1 [Homo sapiens] >gi 32433 hematopoietic proteoglycan core protein (AA 1 - 158) [Homo sapiens] >pir A35183 A28058 secretory granule proteoglycan core prote	gil583141	24	434	98	100	HKMLZ60
422	833207	putative Rab5-interacting protein {clone L1-57} [human, HeLa cells, Peptide Partial, 122 aa] [Homo sapiens] GTP-binding protein [Homo sapiens] >pir G34323 G34323 GTP-binding protein Rab6 - human	gil190420	57	542	81	81	11WAFH33
423	835940		bbs 180090	126	464	78	81	HINFHV44
424	836953		gil550072	388	1038	99	99	HMEFS23

837105		ribosomal protein L5 [Homo sapiens]	gi 550013	860	1168	IILIAS90
837300		>pir S55912 S55912 ribosomal protein L5, cytosolic - human >gil1658578 ribosomal L5 protein [Homo sapiens] {SUB 153-297} Length = 297		276	494	HODHJ94
837373				1	714	HIASC92
837687		protein trafficking protein [Homo sapiens] >gnl PIDe239969 transmembrane protein [Homo sapiens] >gnl PIDe1309760 (AJ004913) integral membrane protein, Tmp21-I (p23) [Homo sapiens] >pir G01159[G01159 protein trafficking protein tmp21-I - human >sp E13097	gil1407826	435	953	HSLBF05
837991		procollagen C-proteinase [Homo sapiens] >sp Q13292 Q13292 PROCOLLAGEN C- PROTEINASE. Length = 986	gil1245357	1	294	HPJCY94
838442		cyclin C [Homo sapiens] >pir A40268 A40268 cyclin C - human >sp P24863 CGIC_HUMAN G1/S-SPECIFIC CYCLIN C. Length = 303 (AF016369) U4/U6 small nuclear ribonucleoprotein hPrp4 [Homo sapiens] >sp O43445 O43445 U4/U6 SMALL NUCLEAR RIBONUCLEOPROTEIN HPRP4. Length = 522	gil1117984	3	506	HAUBJ52
840541				127	549	IHWQA57
840543			gil2708305	40	1020	IHWBEJ29
840550		AZ-1 [Mus musculus] >gnl PIDd1008454 pre- acrosome localization protein [Mus musculus] >pir S63993 S63993 acrosomal protein AZ1 - mouse >sp Q62036 Q62036 S-AZACYTIDINE INDUCED PROTEIN (PRE-ACROSOME LOCALIZATION PROTEIN). Length = 1060	gnl PIDd1019745	1	141	IWBFM54
840563				382	723	HADFY02
840565				1	300	IHGCV14

50

448	840610	plakoglobin [Homo sapiens] >sp Q1S1S1 Q1S1S1 PLAKOGLOBIN. >gnl PID d1010077 plakoglobin [Homo sapiens] {SUB 239-409} Length = 745	gnl PID e214034	1784	2818	94	94	HBCNU40
449	840611	B-IND1 protein [Mus musculus]		657	848			HUFAT62
450	840612	>sp O09003 O09003 B-IND1 PROTEIN. Length = 189	gnl PID e1192419	130	1242	85	86	HWLFV07
451	840615	casein kinase II alpha subunit [Bos taurus] >gil 611 casein kinase alpha subunit [Bos taurus] >gil 177994 casein kinase II alpha subunit [Homo sapiens] >gil 598147 casein kinase II alpha subunit [Homo sapiens] >pir A30319 A30319 casein kinase II (EC 2.7.1.-)	gil 162777	140	1234	94	94	HUKDT16
452	840622	1,4-alpha-glucan branching enzyme [Homo sapiens] >pir A46075 A46075 glycogen branching enzyme - human		135	962			HTXNQ26
453	840623	>sp Q04446 GLGB_HUMAN 1,4-ALPHA- GLUCAN BRANCHING ENZYME (EC 2.4.1.18) (GLYCOGEN BRANCHING ENZYME) (BRANCHER ENZYME). Length = 702	gil 184026	3	542	97	98	HTTEK41
454	840624			1065	1550			HTXBO36

51

52

455	840631	(AL033514) predicted using Genefinder; cDNA EST yk465c10.5 comes from this gene [Caenorhabditis elegans] >sp E1358418 E1358418 Y75B8A.16 PROTEIN. Length = 431	gnl PID e1358418	3	1250	53	73	HTTDDU70
456	840632	(AC004684) putative ribitol dehydrogenase [Arabidopsis thaliana] >sp O80924 O80924 PUTATIVE RIBOTOL DEHYDROGENASE. Length = 321	gi 3236237	2	418	31	50	HTTFY74
457	840633							HTTFA16
458	840634							HTTFG83
459	840635							HTXBW79
460	840636							HTWBE73
461	840637							HTTEZ16
462	840639							HTTET75
463	840640							HTQDA44
464	840650	spermatid perinuclear RNA binding protein [Mus musculus] >pir A57284 A57284 spermatid perinuclear RNA-binding protein Spnr - mouse >sp Q62262 Q62262 SPERMATID PERINUCLEAR RNA-BINDING PROTEIN.	gi 673454	86	940	89	89	HTPAG74
465	840652							HTTCB17

Length = 648

53

466	840653	(AF016507) C-terminal binding protein 2 [Homo sapiens] >sp P56545 CTB2_HUMAN C-TERMINAL BINDING PROTEIN 2. Length = 445	gi 2909777	3	989			HTTDG56
467	840655			1	2139			HTPCP50
468	840659			511	1518	89	89	HTSH154
469	840660	cleavage signal 1 protein [Homo sapiens] >pir JH0629 JH0629 cleavage signal-1 protein - human >sp P28290 CSI_HUMAN CLEAVAGE SIGNAL-1 PROTEIN (CS-1). Length = 249	gi 181123	293	520			HTOJF77
470	840661			3	710			HTLGP71
471	840662			494	1333	90	92	HTOEY44
472	840663	(AF037448) Gry-rbp [Homo sapiens] >sp O60506 O60506 GRY-RBP. Length = 623 complement component C1s [Homo sapiens] >gi 179648 complement subcomponent C1s precursor [Homo sapiens] >gi 763110 complement protein C1s precursor [Homo sapiens] >pir A40496 C1HUS complement subcomponent C1s (EC 3.4.21.42) precursor - human >sp P09871 C1 glypican [Homo sapiens] >pir A36347 A36347 glypican 1 precursor - human >sp P35052 GLYP_HUMAN GLYPICAN-1	gi 3037013	179	466			HTPBY35
473	840670			1132	1647			HTTBJ61
474	840671			210	1001			HTJMJ95
475	840672			3	1739	99	99	HTTIDF09
476	840673		gi 179646	1	690	98	98	HTJAA66
477	840674		gi 31847	208	525	87	87	HTLDZ68

PRECUSOR. Length = 558

478	840677	Similarity to H.influenza ribonuclease PH (SW:RNPH_HAEIN);	237	1010	HTJNE24
479	840678	polynucleotide adenylyltransferase [Bos taurus] >sp P25500 PAP_BOVIN_POLY(A)	3	842	HTGFX11
480	840680	POLYMERASE (EC 2.7.7.19) (PAP) (POLYNUCLEOTIDE	115	555	HTLEI30
481	840691	ADENYLYLTRANSFERASE). (SUB 2-739) Length = 739	1	900	HITEKG75
482	840700	stanniocalcin [Homo sapiens] >gi 975298	54	998	HTELT78
483	840701	stanniocalcin precursor [Homo sapiens] >sp P52823 CSTP_HUMAN STANNIOCALCIN	879	1370	HDQDW52
484	840702	PRECUSOR.	713	955	ITEJY89
485	840705		106	621	ITELU22
486	840715		1	828	HSYBK03
487	840717	(AC005154) similar to protein U28928	561	1058	HSSNA42
488	840718	(PID:g861306) [Homo sapiens] >sp O75223 O75223 WUGSC:11_DJ0777O23.1 PROTEIN. Length = 188	227	562	HSSMV32
489	840719	metallothionein I-F [Homo sapiens] >gi 386866	3	284	HSSNB31
490	840724	human metallothionein-I-F [Homo sapiens] >pir B22634 SMHUIF metallothionein I F - human >sp P04733 MT1F_HUMAN	226	510	HSVBQ73

54

METALLOTHIONEIN-IF (MT-1F). Length = 61

491	840725	Unknown	1259	1501	HSRDA46
492	840727		4	606	HSXCO55
493	840731	apg-2 [Mus musculus] >sp Q61316 HS74_MOUSE HEAT SHOCK 70-RELATED PROTEIN APG-2. Length = 841	22	471	75 HSSAO67
494	840733		3	437	IISGG96
495	840734		228	365	HSRFE65
496	840736	small nuclear ribonucleic protein [Homo sapiens] Length = 92	58	342	100 HSRFE95
497	840737		3	341	HSSFS95
498	840739		196	561	IISLJW05
499	840746	similar to mouse CCL. [Homo sapiens] >sp Q92601 Q92601 MYELOBLAST KIAA0202. Length = 1591	452	1420	85 IISLJL31
500	840748	cytoplasmic antiproteinase, CAP=38 kda intracellular serine proteinase inhibitor [human, placenta, Peptide, 376 aa] [Homo sapiens] Length = 376	65	1441	96 HSRGX11
501	840750	(AC002339) putative ABC transporter [Arabidopsis thaliana] >sp O22950 O22950 ABC TRANSPORTER ISOLOG, 3' PARTIAL (FRAGMENT). Length = 664	507	845	62 HSODA53
502	840751	MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES (EC 2.7.1.117) (MLCK) [CONTAINS: TELOKIN]. Length = 1913	3	2519	90 HTEFV12

503	840757	(AB005624) rig-analog DNA-binding protein [Sus scrofa] >gi 306898 rig-analog protein (putative); putative [Homo sapiens] >gi 337416 human homologue of rat insulinoma gene (rig); putative [Homo sapiens]	gn PI D 1022359	236	568	100	100	100	HKBA1.84
504	840759	transcription factor ZFM1 [Homo sapiens] >sp Q15913 Q15913 TRANSCRIPTION FACTOR ZFM1. Length = 571	gi 100209	481	2073	100	100	100	HSldb36
505	840760	FORMATE ACETYLTRANSFERASE 2 (EC 2.3.1.54) (PYRUVATE FORMATE-LYASE 2) (FRAGMENT). Length = 716	sp D1036490 D1036490	233	529	100	100	100	HSKDG51
506	840770	glyoxalase 1 [Homo sapiens] >gn PI D 1003075 lactoyl glutathione lyase [Homo sapiens] >pir A46714 A46714 lactoylglutathione lyase (EC 4.4.1.5) - human	gi 183258	107	673	99	99	99	HSKCS52
507	840781	(AC003003) Homolog of rat B/K protein product [Homo sapiens] >sp O4330 O4330 HUMAN HOMOLOGUE OF RAT B/K PROTEIN PRODUCT (FRAGMENT). Length = 361	gi 2865208	1	657	93	93	93	HSKHK35
508	840789	polynucleotide phosphorylase (PNPase) [Bacillus subtilis] >gi 184680 polynucleotide phosphorylase [Bacillus subtilis]	gn PI D 1185260	216	347				HIIPSP20
509	840790	>pir S70691 S70691 polynucleotide nucleotidyltransferase (EC 2.7.7.8) alpha chain pnpA - Bacillus subtilis		2	817				HIKSC89
510	840791	>sp P50849 PNPA_BACSU POL (AB001915) NG,NG-dimethylarginine dimethylaminohydrolase [Homo sapiens] Length = 285	gn PI D 1038106	2	493	41	66	66	HHSGD58
511	840798								HIHERQ85
512	840802								HHFES15

513	840803	zinc finger protein [Molgula oculata] >sp Q25473 Q25473 ZINC FINGER PROTEIN. Length = 558	gi 308967	1	1935	36	63	IIIEIRC56
514	840809	(AL022162) dJ454M7.1.1 (Lowe Oculocerebrorenal Syndrome protein OCR-L-1) (isoform 1) [Homo sapiens] >gn P1D1e244699 Lowe oculocerebrorenal syndrome (OCR-L) [Homo sapiens] {SUB 336-813} Length = 813 (AB004903) STAT induced STAT inhibitor-2 [Homo sapiens] >gi 3265033 (AF037989) STAT- induced STAT inhibitor-2 [Homo sapiens] >sp O14508 O14508 STAT INDUCED STAT INHIBITOR-2. Length = 198	gn P1D1e1371023	2	208			IIIEPE84
515	840811			1	690			IIIFBPS1
516	840813			2	214			IIIEMJ45
517	840814			2	154	100	100	HGBIC73
518	840817	Cleavage and Polyadenylation Specificity Factor protein [Bos taurus] >sp P79101 P79101 CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR PROTEIN. Length = 684	gn P1D1e225428	85	864	99	99	IIIEB106
519	840825			2	436			IIIEAB14
520	840826			2022	2360			IIIBFD61
521	840827			14	817			IIIEA1166
522	840828	(AC005757) R32611_2 [Homo sapiens] >sp O75865 O75865 R32611_2 (FRAGMENT). Length = 160 (AF006386) axonemal dynein light chain [Homo sapiens] >sp O14645 O14645 AXONEMAL DYNEIN LIGHT CHAIN. Length = 257	gi 3688090	2	1180	98	99	IIIEAK56
523	840829			130	618			HFVIE96
524	840831			1166	1447			HFXCN75
525	840836			18	566			HFXKK43
526	840837	Length = 257	gi 2352534	322	759	62	80	HGBAG76
527	840838			2	832	100	100	IIEXJP72

528	840841	(AC002333) molybdenum cofactor biosynthesis protein E. isolog [Arabidopsis thaliana]	gi 2281094	2	790	IIIGAMD29
529	840842	>sp O22827 O22827 MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN E. ISOL. OG. Length = 198		216	791	HFPCCK56
530	840843	Gu protein [Homo sapiens] >pir PC6010 PC6010 RNA helicase Gu - human (fragment)		12	791	HFVGM54
531	840845	>sp Q13436 Q13436 NUCLEOLAR RNA HELICASE GU (FRAGMENT). Length = 801		669	1031	IIIGBBY80
532	840847	argininosuccinate lyase [Homo sapiens] >gi 179091 argininosuccinate lyase [Homo sapiens] >pir A31658 WZHURS		151	1044	IIIFPCN94
533	840851	argininosuccinate lyase (EC 4.3.2.1) - human Length = 464		470	2047	IIFOX546
534	840853	(AF064244) intersecin long form [Homo sapiens] >sp G3859855 G3859855 INTERSECTIN LONG FORM. >gi 3859853 (AF064243) intersecin short form [Homo sapiens] {SUB 1-1220} >gi 3930533 (AF064247) intersecin long form [Homo sapiens] {SUB 1209-1263} Length = 172		15	224	HFQXV75
535	840854			149	1183	IIIFBK03
536	840858			249	833	HFPCP42
537	840859		gi 1230564	3	1163	IIIFYQ50
538	840863			1	165	HFIIW33
539	840868		gi 179089	2	1678	HFKEN53
540	840869		gi 3859855	33	632	IIIFKFG36
541	840870			505	831	IIIFKFN13
542	840875			3	617	HFITH86

58

543	840876	(AC004392) Contains similarity to gb U51898 Ca2+-independent phospholipase A2 from Rattus norvegicus. [Arabidopsis thaliana] >sp O80693 O80693 F8K4.6 PROTEIN. Length = 1265	gb 3367519	1	1110	45	70	HF1ZQ25
544	840881	histone H2B.1 [Homo sapiens] >gnl PID e1301465 (AJ223353) Histone H2B [Homo sapiens] >gi 51306 histone H2B-291B (AA 1 - 126) [Mus musculus] >pir S04153 S04153 histone H2B (clone 291B) - mouse >pir F40335 F40335 histone H2B.1 (b) - human >sp E1301465 E1301	gi 184080	3	449	77	77	11F1IR54
545	840883	(AJ000506) Homeodomain protein Meis2c [Mus musculus] >sp P97367 MEI2_MOUSE HOMEOBOX PROTEIN MEIS2 (MEIS1- RELATED PROTEIN 1). Length = 477	gnl PID e330082	3	428	90	90	11F1IIA80 11HPDW66
546	840886			71	964			
547	840887	RNA polymerase I subunit A12.2		1202	1600			11F1IR82
548	840891	[Saccharomyces cerevisiae] >gi 1019685 ORF YJR063w [Saccharomyces cerevisiae] >gi 531231 RNA polymerase I A12.2 subunit [Saccharomyces cerevisiae] >gi 1015737 ORF YJR063w [Saccharomyces cerevisiae] >pir A48107 A48107 DNA-dir histone H2B [Homo sapiens] >pir 137445 137445 histone H2B.1 - human >sp P33778 H2B0_HUMAN HISTONE H2B.1. {SUB 2-126} Length = 126	gi 172462	250	375	64	86	11FCBQ77
549	840892		gi 31977	3	410	98	98	11FEBK16

550	840894	(AF002697) EIB 19K/Bcl-2-binding protein Nip3 [Homo sapiens] >sp O14620 O14620 EIB 19K/BCL-2-BINDING PROTEIN NIP3. Length = 194	gi 2511529	1	705	80	80	HFIIIO60
551	840896	Cdc73p [Saccharomyces cerevisiae] >pir SS9383 SS9383 probable membrane protein YLR418c - yeast (Saccharomyces cerevisiae) >sp Q06697 Q06697 CHROMOSOME XII COSMID 9931. Length = 393	gi 632679	425	1249	28	57	HFIAL02
552	840897	syntaxin-4 [Homo sapiens] >gnl PID e332032 (AJ000541) syntaxin 4 precursor [Homo sapiens] >gi 2570870 (AF026007) syntaxin 4 [Homo sapiens] >pir S52726 S52726 syntaxin-4 - human Length = 297	gi 758105	3	1142	100	100	HFIAW49
553	840898	DNA fragmentation factor-45 [Homo sapiens] >sp O0273 DF45_HUMAN DNA FRAGMENTATION FACTOR-45 (DF45). Length = 331	gi 2065561	3	1100	95	95	HETBS69
554	840904							
555	840905							
556	840908	KIAA0156 gene product is related to Xenopus nucleolin. [Homo sapiens] >sp Q15020 Q15020 ORF. Length = 963	gnl PID d1010577	348	2081	87	87	HETCI63
557	840909	3-methyl-adenine DNA glycosylase [Homo sapiens] Length = 298	gnl PID e224269	2	949	94	94	HIEQAN83
558	840910	MAL protein [Homo sapiens] >gi 435478 MAL-a gene product [Homo sapiens] >gnl PID e1192240 MAL. [Homo sapiens] >pir A29472 A29472 T-cell surface glycoprotein MAL, splice form a - human	gi 307157	1	432	86	93	HETJW92
559	840912							
560	840916							

60

561	840917	(AF020038) NADP-dependent isocitrate dehydrogenase [Homo sapiens] >gi 3641398	518	886	99	99	11ETI212
562	840918	(AF020038) NADP-dependent isocitrate dehydrogenase [Homo sapien	231	1508			HAJCO38
563	840922		839	1033			11ELGB82
564	840923		1044	1289			11EQAN39
565	840927		119	364			11EMFU44
566	840928		2	1258			11EMCG01
567	840929	helix-loop-helix phosphoprotein [Homo sapiens] >gi 292055 helix-loop-helix phosphoprotein [Homo sapiens] >pir 53020 53020 G-0/G-1 switch regulatory protein 8 - human >pir 65984 65984 helix-loop-helix phosphoprotein - human Length = 211 (AF002282) alpha-actinin-2 associated LIM protein [Homo sapiens] >sp O60440 O60440 ALPHA-ACTININ-2 ASSOCIATED LIM PROTEIN. Length = 316 similar to thiolesterase;	3	662	92	92	11EOMQ95
568	840930		3	1019	99	99	11EGAD28
569	840931		1	1164	49	67	11EMFC70
570	840941		2	781			11EGAL15
571	840944	cofactor E [Homo sapiens] >sp Q15813 Q15813 COFACTOR E. Length = 527	822	1685	98	98	11ELFC44
572	840945	lanosterol synthase [human, fetal liver, Peptide, 732 aa] [Homo sapiens] >gnl PID d1010523	1067	1435	99	100	11EAS77
573	840948	lanosterol synthase [Homo sapiens] >gi 951314 2,3-oxidosqualene-lanosterol cyclase [Homo sapiens] >pir JC4194 JC4194 lanosterol synthase (EC 5.4.99.7) - human >sp p	3	326			11E9ST22

574	840949	(AJ005324) glutamate permease [synthetic construct] >gnl PID e1360147 (AJ005327) glutamate permease [synthetic construct] >gnl PID e1360153 (AJ005330) glutamate permease [synthetic construct] Length = 459 P43=mitochondrial elongation factor homolog [human, liver, Peptide, 452 aa] [Homo sapiens] >pir I53499 I53499 translation elongation factor TU-like protein P43, mitochondrial - human Length = 452 RNase L inhibitor (clone 8) - human Length = 599 FUSE binding protein 2 [Homo sapiens] >sp Q92945 Q92945 FUSE BINDING PROTEIN 2 (FRAGMENT). Length = 652 phosphomannose isomerase [Homo sapiens] >pir S41122 S41122 mannose-6-phosphate isomerase (EC 5.3.1.8) - human >sp P34949 MANA_HUMAN MANNOSE-6-PHOSPHATE ISOMERASE (EC 5.3.1.8) (PHOSPHOMANNOSE ISOMERASE) (PMI) (PHOSPHOHEXOMUTASE). {SUB 2-423} Length = 423	gnl PID e1360141	3	101	95	95	IIE9RM92
575	840953		bbs I60014	1	1437	100	100	IIE1GIM94
576	840954		pir S63672 S63672	69	1949	95	95	IIE9HIC20
577	840958		g i I575607	154	465	57	58	HFLVB33
578	840960		g i I6017	224	670	100	100	HIEAD70
579	840968			375	2222			IIEBFI129
580	840969			1054	1530			IIE9PIB53
581	840972			1	387			IIE8UII14
582	840973			548	874			IIE9DI168
583	840975			1	159			IIE9GO90
584	840978			1433	1765			IIE9NG78

585	840980	nerve growth factor [Homo sapiens] >gi 32031 pleiotrophin [Homo sapiens] >bbs 19887 pleiotrophin, PTN [human, Peptide, 168 aa] [Homo sapiens] >bbs 130735 heparin-binding neurite outgrowth promoting factor, HBNF {alternatively spliced} [human, Peptide, 16	gij 183890	75	833	90	90	HEBFE14
586	840982			81	359			HE8ES49
587	840985			3	830			HE8UK50
588	840989	(AB016247) sterol-C5-desaturase [Homo sapiens] >sp O75845 O75845 STEROL-C5- DESATURASE (EC 1.3.3.2) (LATHOSTEROL OXIDASE). Length = 299	gnl PID d1034698	107	1027	99	100	HE8FM74
589	840991	>sp O43524 O43524 FORKHEAD PROTEIN. Length = 673	gij 2895494	861	1559	81	81	HE8FA09
590	840996	ATP:citrate lyase [Homo sapiens] >sp Q13037 Q13037 ATP:CITRATE LYASE. Length = 1101	gij 603074	818	1906	99	99	HE8MY23
591	840997	LIV-1 protein [Homo sapiens] >pir G02273 G02273 LIV-1 protein - human >sp Q13433 Q13433 ESTROGEN REGULATED LIV-1 PROTEIN. Length = 752	gij 256001	3	1193	75	75	HE8DR57
592	840998			1	390			HE2BN26
593	840999			855	1013			HE8DJ30
594	841000			1	279			HE6DC57
595	841002			363	812			HE8BT63
596	841003			94	315			HE2DX28
597	841008	Aop1_Human, MER5(Aop1_Mouse)-like protein [Homo sapiens] >gij 854126 humer [Homo sapiens] {SUB 227-256} Length = 256	gnl PID d1008985	1	672	99	99	HE8AU49

598	841013	(AB011004) UDP-N-acetylglucosamine pyrophosphorylase [Homo sapiens] >sp Q16222 Q16222 AGX-1 ANTIGEN (FRAGMENT). Length = 505	gnl P1D d1032151	265	1836	99	99	IIDT'AU64
599	841014	fumarase precursor [Homo sapiens] >gi 4097195 fumarase [Homo sapiens] >sp P07954 FUMH_HUMAN FUMARATE HYDRATASE, MITOCHONDRIAL PRECURSOR (EC 4.2.1.2) (FUMARASE). >sp G4097195 G4097195 FUMARASE (EC 4.2.1.2). Length = 510	gi 1545996	178	1185	96	96	IIE2EB32
600	841015	Ran [Canis familiaris] >gi 190879 ras-like protein [Homo sapiens] >gi 2967848 (AF052578) androgen receptor associated protein 24 [Homo sapiens] >gi 727167 Ran [Mus musculus] >bbs 180269 GTP-binding protein [mice, C3H/HeJ spleens, LDS responder, Peptide, 2	gi 924	48	425	100	100	IIE2DT31
601	841018			1	150			HE2EA79
602	841019			94	228			HDTGC76
603	841024			34	750			HE9CO25
604	841025	Id-2H [Homo sapiens] >pir A40227 A40227 transcription repressor Id-2 - human >sp Q02363 ID2_HUMAN DNA-BINDING PROTEIN INHIBITOR ID-2. Length = 134	gnl P1D d1003496	75	401	100	100	IIDTIDZ04
605	841026			3	599			HDTGP42
606	841027			1	489			HDRMB48
607	841029			1	528			IIDTAG94
608	841030			515	721			IIDTCK45
609	841031			23	145			IIDSAL27

64

610	841034	G-rich sequence factor-1 [Homo sapiens] >gi 517196 G-rich sequence factor-1 [Homo sapiens] >sp Q12849 GRF1_HUMAN G-RICH SEQUENCE FACTOR-1 (GRSF-1). >pir S48081 S48081 GRSF-1 protein - human (fragment) {SUB 94-424} Length = 424	gi 517196	267	449	95	98	IIDQDI160
611	841036	(AC002340) putative RNA helicase A, 5' partial		1201	1542			IIDPTM31
612	841039	[Arabidopsis thaliana] >sp O49345 O49345 PUTATIVE RNA HELICASE A, 5' PARTIAL (FRAGMENT). Length = 1114	gi 2880057	763	2112	60	76	IIDQFB71
613	841040	(AF071202) ABC transporter MOAT-B [Homo sapiens] >sp G3335173 G3335173 ABC TRANSPORTER MOAT-B. Length = 1325	gi 3335173	2	1339	92	92	IIDQDF77
614	841048	(AC003682) ZNF134 [Homo sapiens]		1	1338			IIDPXU60
615	841049	>sp G2689444 G2689444 ZNF134. Length = 427	gi 2689444	3	347	97	97	IIDPKK77
616	841050	monoamine oxidase A [Homo sapiens]		705	947			IIDPUP64
617	841052	>gi 187353 monoamine oxidase A [Homo sapiens] >gi 187355 monoamine oxidase A [Homo sapiens] >pir A36175 A36175 amine oxidase (flavin-containing) (EC 1.4.3.4) A - human >sp P21397 AOFA_HUMAN AMINE OXIDASE [FLAVIN-CONTAINI	gi 187351	1	1194	95	95	IIDPRJ46
618	841054			60	1262			IIDPXL80
619	841055			23	346			IIDPMK92
620	841056			492	695			IIDPVB33
621	841060			612	851			IIDPXB24

65

622	841061	quinone oxidoreductase [Homo sapiens] >gi 516534 quinone oxidoreductase2 [Homo sapiens] >pir A32667 A32667 NAD(P)H dehydrogenase (quinone) (EC 1.6.99.2) 2 - human Length = 231	gi 190818	21	614	100	100	IIPBQ60
623	841062	histone deacetylase HD1 [Homo sapiens] >sp Q13547 HDA1_HUMAN HISTONE DEACETYLASE 1 (HD1). Length = 482 (A1.009194) SWISS-PROT:P38861;	gi 1277084	67	1530	90	90	IIDPPA96
624	841063	NONSENSE-MEDIATED MRNA DECAY PROTEIN 3.; SACCHAROMYCES CEREVISIAE mannosyl-oligosaccharide 1,2-alpha-mannosidase (EC 3.2.1.113) - rabbit (fragment) >gi 474282 mannosyl-oligosaccharide alpha-1,2-mannosidase [Oryctolagus cuniculus] {SUB 12-480} Length = 480	gnl PID e1251068	2	592	69	82	IIDPJQ57
625	841067	mannosyl-oligosaccharide 1,2-alpha-mannosidase (EC 3.2.1.113) - rabbit (fragment) >gi 474282 mannosyl-oligosaccharide alpha-1,2-mannosidase [Oryctolagus cuniculus] {SUB 12-480} Length = 480	pir B54408 B54408	2	592	59	83	IIDPQE64
626	841074	14-3-3 protein [Homo sapiens] >gi 32464 HS1 gene product [Homo sapiens] >pir S15076 S15076 protein kinase regulator 14.3.3 - human >sp P27348 I43T_HUMAN 14-3-3 PROTEIN TAU (14-3-3 PROTEIN THETA) (14-3-3 PROTEIN T-CELL) (HS1 PROTEIN). >gi 3387922 (AF070556	gi 23222	188	907	98	99	HE8NS76
627	841076	(AE000715) ribosomal protein L20 [Aquifex aeolicus] >pir C70382 C70382 ribosomal protein L20 - Aquifex aeolicus >sp O67086 O67086 50S RIBOSOMAL PROTEIN L20. Length = 118	gi 2983472	96	755	41	65	IIDPMG95
628	841081			2	541			IIDPQC09
629	841083			1	480			HIDPCX80
630	841089			321	551			HDPND16

631	841093	(AF035646) Rab10 [Mus musculus] >sp O88386 O88386 RAB10. Length = 200	gi 3406428	479	1132	100	100	100	IIDPP129
632	841097	(AF090867) guanosine monophosphate reductase [Rattus norvegicus] >sp G3907579 G3907579 GUANOSINE MONOPHOSPHATE REDUCTASE. Length = 345	gi 3907579	267	1061	78	90	90	IIDPF1378
633	841098	GATA-binding protein [Homo sapiens] >pir A40815 A40815 transcription factor GATA- 2 (version 1) - human >sp P23769 GAT2_HUMAN ENDOTHELIAL TRANSCRIPTION FACTOR GATA-2. Length = 480	gi 182996	1	384	90	91	91	IIDABX64
634	841101	phosphatidylcholine transfer protein [Bos taurus] >pir A91092 EPBO phosphatidylcholine transfer protein - bovine >sp P02720 PPCT_BOVIN PHOSPHATIDYLCHOLINE TRANSFER PROTEIN (PC-TP). Length = 213	gi 710419	3	1004	35	55	55	IIDPBQ32
635	841113	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (hpcE) [Methanococcus jannaschii] >pir F64506 F64506 2-hydroxyhepta-2,4-diene- 1,7-dioate isomerase homolog - Methanococcus jannaschii >sp Q59050 Q59050 HYPOTHETICAL PROTEIN MJ1656. Length = 237	gi 1500558	133	1137	50	74	74	IIDBA185
636	841115			58	396				IIDLAZ62
637	841116			47	682				HDPBJ61
638	841117			1	1179				HDFMB93
639	841125			1	117				HICYB178
640	841127			2	859				IIDABQ85

641	841128	collagenase stimulatory factor [Homo sapiens] >gi 1209374 amino acid feature: intracellular domain, aa 707 .. 829; amino acid feature: transmembrane domain, aa 638 .. 706; amino acid feature: extracellular domain, aa 86 .. 637 [Homo sapiens] >gi 34449 M6	gi 409357	64	891	100	100	100	IIDPFI118
642	841132	myosin-I, Myr 1c (alternatively spliced) - rat Length = 1078	pir B45439 B45439	1	1428				IIDPFI170
643	841133			4	1710	91	89		HCYBI.17
644	841134	gamma SNAP [Homo sapiens] Length = 312 homologous to mouse gene PC326: GenBank Accession Number M95564 [Homo sapiens] >sp Q12839 Q12839 (H326). Length = 597	gi 1685288 gi 458692	2	802	100	100		IIDAAC32
645	841135			124	765	81	81		IIDABE30
646	841136	imogen 38 [Homo sapiens] >sp Q92665 Q92665 IMOGEN 38. Length = 395	gh PID e218584	514	735				IICQDF95
647	841138			3	1238	80	80		IIDABK25
648	841139	(AF038957) translation initiation factor 4e [Homo sapiens] >sp O75349 O75349 TRANSLATION INITIATION FACTOR 4E. Length = 236	gi 3329384	347	478				IICQBI160
649	841141			192	833				IIDPBQ85
650	841142			452	1051				HCQAM05
651	841145			1022	1366				HCNSQ35
652	841146			864	1061				HCMSW06
653	841150			115	387	86	83		IICQAG10

654	841153	argininosuccinate synthetase [Homo sapiens] >gi 28872 argininosuccinate synthetase (aa 1-412) [Homo sapiens] >pir A01195 AJHURS argininosuccinate synthase (EC 6.3.4.5) - human >sp P00966 ASSY_HUMAN ARGINOSUCCINATE SYNTHASE (EC 6.3.4.5) (CITRULLINE--ASPA (AF084260) signalosome subunit 2 [Homo sapiens] >gi 3639069 (AF087688) alien-like protein [Mus musculus] >sp O88950 O88950 ALIEN-LIKE PROTEIN. >sp G3514097 G3514097 SIGNALOSOME SUBUNIT 2. >gi 3309166 (AF071312) COP9 complex subunit 2 [Mus musculus] {SUB 4 carcinoma-associated antigen GA733-2 [Homo sapiens] >gi 182906 carcinoma-associated antigen GA733-2 [Homo sapiens] >pir B48149 B48149 epithelial glycoprotein antigen GA733-2 precursor - human Length = 314	gi 179057	1207	2532	96	96	HCYBC10
655	841154	(AF084260) signalosome subunit 2 [Homo sapiens] >gi 3639069 (AF087688) alien-like protein [Mus musculus] >sp O88950 O88950 ALIEN-LIKE PROTEIN. >sp G3514097 G3514097 SIGNALOSOME SUBUNIT 2. >gi 3309166 (AF071312) COP9 complex subunit 2 [Mus musculus] {SUB 4 carcinoma-associated antigen GA733-2 [Homo sapiens] >gi 182906 carcinoma-associated antigen GA733-2 [Homo sapiens] >pir B48149 B48149 epithelial glycoprotein antigen GA733-2 precursor - human Length = 314	gi 3514097	1	1368	100	100	HCMSB29
656	841156	(AF084260) signalosome subunit 2 [Homo sapiens] >gi 3639069 (AF087688) alien-like protein [Mus musculus] >sp O88950 O88950 ALIEN-LIKE PROTEIN. >sp G3514097 G3514097 SIGNALOSOME SUBUNIT 2. >gi 3309166 (AF071312) COP9 complex subunit 2 [Mus musculus] {SUB 4 carcinoma-associated antigen GA733-2 [Homo sapiens] >gi 182906 carcinoma-associated antigen GA733-2 [Homo sapiens] >pir B48149 B48149 epithelial glycoprotein antigen GA733-2 precursor - human Length = 314	gi 182896	6	1130	86	86	HCIAA60
657	841157	collagen pro-alpha-1 type I chain [Mus musculus] >pir S57243 S21626 collagen alpha 1(I) chain precursor - mouse >sp P11087 CA11_MOUSE PROCOLLAGEN ALPHA 1(I) CHAIN PRECURSOR. >gi 192262 pro-alpha-1 type I collagen [Mus musculus] {SUB 518-1128} >gi 192264 p	gi 470674	88	336	36	42	HCICJ07
658	841159			510	818			HCICK84
659	841164			2	463			HCIAZ66
660	841167			982	1305			HCIOG20

661	841170	SRp30c [Homo sapiens] >gnl PID e 248292 (AL021546) pre-mRNA splicing factor SRp30c [Homo sapiens] >gi 4099429 splicing factor SRp30c [Homo sapiens] >pir S59075 S59075 splicing factor SRp30c - human >sp G4099429 G4099429 SPLICING FACTOR SRp30C. Length = 22	gi 1049078	2	760	81	81	HCHOE21
662	841173	spermidine synthase [Homo sapiens] >pir A32610 A32610 spermidine synthase (EC 2.5.1.16) - human Length = 302	gi 338394	2	931	97	97	IICHBQ07
663	841176	thyroid receptor interactor [Homo sapiens] Length = 152		561	683			IICFQI36
664	841178	(AF029777) hGCN5 [Homo sapiens] >sp G3220164 G3220164 HGCN5. >gi 1491935 histone acetyltransferase [Homo sapiens] {SUB 362-837} >sp G1911495 G1911495 HGCN5=TRANSCRIPTIONAL ADAPTOR. {SUB 411-837} Length = 837	gi 703110	65	460	99	100	HCGBQ34
665	841180		gi 3220164	553	1530	97	97	HCGLC82
666	841181			2	283			HCFMN22
667	841182	70 K protein (AA 1-614) [Homo sapiens] >pir A25707 A25707 U1 snRNP 70K protein - human >gi 337447 small ribonucleoprotein 70 kd protein [Homo sapiens] {SUB 178-614} >gi 602021 hU1-70K protein (302 AA) [Homo sapiens] {SUB 227-527} Length = 614	gi 36100	251	988	100	100	HCFNJ56
668	841185			342	536			IICFNF67
669	841187			458	1096			HCGAA74
670	841188	DNA repair endonuclease subunit [Homo sapiens] Length = 905	gi 1524411	2	2749	92	92	IICFMK76
671	841189			336	926			IICFMC34

71

672	841192	methylmalonyl-CoA mutase [Homo sapiens] >sp P22033 MUTA_HUMAN METHYLMALONYL-COA MUTASE PRECURSOR (EC 5.4.99.2) (MCM). Length = 750	gi 187452	1	1428	99	99	IICFMO54
673	841194	(AF039405) arsenite-translocating ATPase [Mus musculus] >sp O54984 O54984 ARSENITE- TRANSLOCATING ATPASE. Length = 350	gi 2745900	182	1138	95	95	IICGAB52
674	841195	(AF015037) endooligopeptidase A related protein; EOPA related protein [Oryctolagus cuniculus] >sp O46480 O46480 ENDOOLIGOPEPTIDASE A RELATED PROTEIN (FRAGMENT). Length = 667	gi 2827886	3	623			IICFWM29
675	841198			2	913			IICFBC32
676	841200			35	703	75	81	IICFER84
677	841201	rhoB [Homo sapiens] >gi 206656 rhoB [Rattus norvegicus] >gnl PID e258480 RHOB [Mus musculus] >pir A01372 TVHURH GTP-binding protein rhoB - human >pir A39727 TVKTRH GTP-binding protein rhoB - rat >pir C5075 JC5075 GTP-binding protein rhoB - mouse >gi 3373	gi 36032	158	571	100	100	IICFBD63
678	841202			66	1229	100	100	IICFHOV21
679	841209	PTB-associated splicing factor [Homo sapiens] >pir A46302 A46302 PTB-associated splicing factor, long form - human >gi 23712 myoblast antigen 24.1D5 [Homo sapiens] (SUB 312-707) >gi 4063717 (AF110499) PTB-associated splicing factor [Mus musculus] (SUB 377	gi 38458	1	552			IICDMF27
680	841210			2	1405	93	93	IICFMT64

681	841213	G9a [Homo sapiens] >pir S30385 S30385 G9a protein - human >sp Q14349 Q14349 G9A PROTEIN CONTAINING ANKYRIN-LIKE REPEATS. Length = 1001	gi 287865	3	344	82	84	ICIEFE38
682	841217			2	1198			HCEIV79
683	841219	SMOOTH MUSCLE MYOSIN HEAVY CHAIN (FRAGMENT). Length = 1052	sp D1037960 D1037960	208	774	95	97	IIBZSI02
684	841222			29	856			IICDCI63
685	841223			2088	2486			HCEBW38
686	841224	RNA polymerase II elongation factor ELL2 [Homo sapiens] >sp O00472 ELL2_HUMAN RNA POLYMERASE II ELONGATION FACTOR ELL2. Length = 640	gi 1946347	2	2032	95	95	HCE2DI5
687	841226			2	373			IICCMD50
688	841227			1	831			HBZAK55
689	841228	F25H9.7 [Caenorhabditis elegans] >gnl PIDe 1346003 F25H9.7 [Caenorhabditis elegans] >sp P91989 P91989 F25H9.7 PROTEIN. Length = 154	gnl PIDe 1346003	3	407	46	62	HCDEA07
690	841231			279	977			IIBXCC66

691	841232	MHC HLA-RD protein [Homo sapiens] >pir A33640 A33640 class III histocompatibility antigen RD - human Length = 382	gi 386949	3	461	94	95	11C'EIS91
692	841233	(AF069984) nitrilase homolog 1 [Homo sapiens] >gi 3228666 (AF069987) nitrilase 1 [Homo sapiens] >sp O76091 O76091 NITRILASE HOMOLOG 1. Length = 327	gi 3242978	2	673	94	95	HBUAF56
693	841234	(AJ005073) Alix [Mus musculus] >sp O88695 O88695 ALIX. Length = 869	gnl PID e1318710	561	2564	89	91	11BWC170
694	841236			187	483			11BXG1B85
695	841238			168	389			11BXFF92
696	841239			405	605			11BMU108
697	841242			169	360			11BNAT03
698	841243			3	281			11BMTQ45
699	841248	phorbol 3 [Homo sapiens] >sp G4097433 G4097433 PHORBOLIN 3. Length = 235	gi 4097433	3	668	46	62	11BUAC02
700	841250			2	1309			11BJEC31
701	841251			5	247			11BJLL24
702	841254			879	1136			11BZSH07
703	841263			1	354			11BJDS57
704	841266			182	337			11BJFN11
705	841269	(AL021958) fadE9 [Mycobacterium tuberculosis] >sp O53815 O53815 ACYL-COA DEHYDROGENASE. Length = 390	gnl PID e1253290	93	1130	51	70	11BIDAC79
706	841272	p67 myc protein [Homo sapiens] >sp D1001846 D1001846 P67 MYC PROTEIN (FRAGMENT). Length = 454	gnl PID d1001846	20	622	100	100	11BJFJ36
707	841273			697	948			11BFMD57
708	841276			244	423			11BNAE62

73

709	841277	NADH-UBIQUINONE OXIDOREDUCTASE 39 KD SUBUNIT PRECURSOR (EC 1.6.5.3) (EC 1.6.99.3) (COMPLEX I-39KD) (CI-39KD). >gil189049 NADH dehydrogenase (ubiquinone) [Homo sapiens] {SUB 3-377} Length = 377 gag polyprotein - human endogenous virus S71 Length = 608	sp Q16795 NUEM_HUMA N	2	1171	94	94	HBICG75
710	841278		pir A46312 A46312	119	415	44	56	HAATDB46
711	841279			187	645			IPIAF81
712	841280	(AF061513) candidate adaptor protein CED-6 [Caenorhabditis elegans] >sp O76337 O76337 CANDIDATE ADAPTOR PROTEIN CED-6. Length = 492	gil3253308	888	1823	50	69	HBCAS37
713	841282			219	368			HAATAM48
714	841283			2530	2880			HBAFS89
715	841286	(AC003096) putative protein phosphatase 2C [Arabidopsis thaliana] >sp O64583 O64583 HYPOTHETICAL 26.4 KD PROTEIN. Length = 239	gil3132471	201	1319	57	80	HAHCP59
716	841287			3	248			HARMV18
717	841288	(AL021428) hypothetical protein Rv0068 [Mycobacterium tuberculosis] >sp O53613 O53613 OXIDOREDUCTASE. Length = 303	gnl PIDe1245998	3	821			HARMM85
718	841291	selenoprotein P [Homo sapiens] Length = 381		293	1012	88	89	HBMCL13
719	841292	SSR gamma subunit [Rattus norvegicus] >pir S33294 S33294 translocon-associated protein gamma chain - rat Length = 185 microtubule associated protein [Homo sapiens] >pir J37356 J37356 epithelial microtubule- associated protein, 115K - human >sp Q14244 Q14244 MICROTUBULE ASSOCIATED PROTEIN. Length = 749	gil312702	2	664	98	98	HARAIS2
720	841294		gil414115	3	1265	99	99	HAAPOR25

74

721	841296	protein disulfide isomerase-related protein [Homo sapiens] >pir A23723 A23723 protein disulfide-isomerase (EC 5.3.4.1) ERp72 precursor - human >sp P13667 ER72_HUMAN PROTEIN DISULFIDE ISOMERASE-RELATED PROTEIN PRECURSOR (ERP72). Length = 645 Gps1 [Homo sapiens] >pir G01646 G01646 Gps1 - human >sp Q13098 GPS1_HUMAN G PROTEIN PATHWAY SUPPRESSOR 1 (GPS1 PROTEIN) (MFH PROTEIN). {SUB 30-500} Length = 500	gj 181508	2	1405	96	96	HASAS34
722	841298		gj 644879	3	1067	91	91	IIATTA149
723	841301	synexin [Homo sapiens]		10	231			IIAPNO69
724	841303	>sp P20073 ANX7_HUMAN ANNEXIN VII (SYNEXIN). Length = 466 (AB000199) CCA2 protein [Rattus norvegicus] >sp O35048 O35048 CCA2 PROTEIN. Length = 338	gj 338244	3	1457	100	100	HAOMG39
725	841304		dbj 1AB000199_1	3	707	89	89	IIATPA1E10
726	841305	similar to RNA binding protein;		399	1274			IIAMHD70
727	841309	>sp Q19706 IF35_CABEL PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 RNA-BINDING SUBUNIT (EIF-3 RNA-BINDING SUBUNIT) (EIF3 P33) (TRANSLATION INITIATION F	gnl PIDe1345859	137	1699	48	63	IIAPAJ60
728	841314	(AJ224819) tumor suppressor [Homo sapiens]		3	920			IIAMGN09
729	841316	>sp O60858 O60858 TUMOR SUPPRESSOR. Length = 407	gnl PIDe1292742	185	1420	93	93	HAJCP55

75

730	841318	replication control protein 1 [Homo sapiens] >pir G02329 G02329 replication control protein 1 - human >sp Q13471 Q13471 REPLICATION CONTROL PROTEIN 1. Length = 861	gij1171204	170	436	100	100	100	IIAMFQ80
731	841321	hnRNP A2 protein [Homo sapiens] >gij P1D d1006583 hnRNP A2 protein [Homo sapiens] >gij 500638 hnRNP protein A2 [Homo sapiens] Length = 341	gij337449	3	656	100	100	100	HBJMK69
732	841324	chimeric IFNalpha/beta-receptor [Homo sapiens] >gij306914 interferon-alpha receptor precursor [Homo sapiens] >pir A32694 A32694 interferon alpha receptor precursor - human >sp P17181 IINR1_HUMAN INTERFERON- ALPHA/BETA RECEPTOR ALPHA CHAIN PRECURSOR (IFN-ALP	gij P1D e251628	31	1755	99	99	99	IIAMGIF04
733	841326	Rch1 [Homo sapiens] >gij899539 hSRP1alpha [Homo sapiens] >pir A56516 A56516 nuclear localization sequence receptor SRP1 alpha - human >sp P52292 HMA2_HUMAN IMPORTIN ALPHA-2 SUBUNIT (KARYOPHERIN ALPHA-2 SUBUNIT) (SRP1-ALPHA) (RAG COHORT PROTEIN 1). Length	gij791185	3	1715	97	97	97	IIAMFV20
734	841328	nuclear ribonucleoprotein [Homo sapiens] >gij35772 polypirimidine tract binding protein [Homo sapiens] >pir S26294 S26294 polypirimidine tract-binding protein - human Length = 557	gij32354	2	1126	89	89	89	IIAMGIF52
735	841329	box protein p72 [Homo sapiens] >pir S72367 S72367 ATP-dependent RNA helicase - human >sp Q92841 P72_HUMAN PROBABLE RNA-DEPENDENT HELICASE	gij P1D e1249592	93	671	100	100	100	IIAJBV54

P72 (DEAD-BOX PROTEIN P72). Length = 650

1/23/2006, EAST Version: 2.0.1.4

736	841330	(AF002228) tbx3 [Homo sapiens] >sp O15119 O15119.TBX3 (FRAGMENT). Length = 468	gi 3041821	3	1097	91	91	HAAJAZ71
737	841333	(AB010882) hSNF2H [Homo sapiens] >sp O60264 O60264.HSNF2H. Length = 1052	gnl P1D d1026101	1	2004	92	92	HAAJBA64
738	841334	SDF2 [Mus musculus] >pir C5105 C5105 stromal cell-derived factor 2 - mouse >sp P97307 P97307.STROMAL CELL DERIVED FACTOR 2 (SDF2). Length = 211	gnl P1D d1009954	3	713	59	71	HAAJBI68
739	841335			443	946			HAAJA172
740	841336			1	1557			HAAJCD33
741	841337			263	1375			HAAJAO95
742	841339	transcription factor SC1 [Homo sapiens] >sp Q13176 Q13176.TRANSCRIPTION FACTOR SC1. Length = 359	gi 833833	27	740	88	89	HAAJCB95
743	841340			820	1017			HAAJAD20
744	841341			3	359			HAAJAI18
745	841342			1145	1417			HAAJAI64
746	841343	cellular nucleic acid binding protein [Mus musculus] >pir I49259 I49259 cellular nucleic acid binding protein - mouse Length = 178 (AF038844) MKP-1 like protein tyrosine phosphatase [Homo sapiens] >sp G4104681 G4104681 MKP-1 LIKE PROTEIN TYROSINE PHOSPHATASE. Length = 198	gi 854675	263	685	100	100	HAAJMG35
747	841347		gi 4104681	161	409	100	100	HAAHSE21

748	841352	ribosomal protein L35 [Homo sapiens] >pir G01477 G01477 ribosomal protein L35 - human Length = 123	gi 562074	3	461	100	100	100	IIBJF14
749	841353	FKBP65 binding protein [Mus musculus] >pir 49669 49669 FKBP65 binding protein - mouse >sp Q61576 Q61576 FK506 BINDING PROTEIN 6 (65 KDA) (FKBP65 BINDING PROTEIN). Length = 581	gi 894162	73	462	92	96	96	IHAIC069
750	841354			115	630				IAPNQ64
751	841360			1	816				IAMI:M60
752	841366			222	1319				IAMGA45
753	841405	cathepsin O [Homo sapiens] >gi 562757 Cathepsin O [Homo sapiens] >bbs 172248 cathepsin O2 [human, spleen. Peptide, 329 aa] [Homo sapiens] >pir JC2476 JC2476 cathepsin K (EC 3.4.22.-) precursor - human signal recognition particle receptor beta subunit [Mus musculus] >pir A56487 A56487 signal recognition particle receptor beta chain - mouse Length = 269	gi 606923	24	1106	100	100	100	HOABW85
754	841526	DNA-binding protein [Homo sapiens] >pir S69501 S69501 DNA-binding protein A variant - human >sp Q14121 Q14121 DNA- BINDING PROTEIN. Length = 372 mitochondrial ATPase inhibitor [Rattus norvegicus] >gnl PID d1002924 ATPase inhibitor protein precursor [Rattus sp.] >pir JS0738 JS0738 ATPase inhibitor protein precursor, mitochondrial - rat >sp Q03344 IATP_RAT ATPASE INHIBITOR, MITOCHONDRIAL	gi 600886	3	848	86	88	88	IHA3AD39
755	841712			3	698	76	76	76	IIBJ193
756	841860			1984	2352				IPIAP58
757	842042			2	817				IIBMXV50
758	842453		gi 517226	13	276	76	88	88	IIBKDV52

78

79

PRECURSOR.

759	842635	(AF010313) Pig8 [Homo sapiens]	gi 2415302	268	936	HPIII120
760	842927	>sp O14681 O14681_PIG8. Length = 318		2	1630	HCE3G66
761	842988	(AF010187) FGF-1 intracellular binding protein		940	1152	IIOSAI376
762	843080	[Homo sapiens] >gi 2738522 (AF010188) FGF-1		2050	2442	IIDPBA08
763	843237	intracellular binding protein [Cercopithecus		370	1359	IIEIJJ27
764	843381	aethiops] >gi 2738520 (AF010187) FGF-1		520	777	HSIGN74
765	843718	intracellular binding protein [Homo sapiens]		212	262	HMEGI84
766	843823	>gi 2738522 (AF010188) FGF-1 intrac		2	1414	IIIESE185
767	844056	(AF010187) FGF-1 intracellular binding protein	gi 2738520	2	751	IIIE8U238
768	844325	[Homo sapiens] actin binding protein MAYVEN	gi 3789797	46	1056	HPRSB90
769	844344	(AF059569) actin binding protein MAYVEN		1	303	IIBJNC37
770	844368	[Homo sapiens] >sp G3789797 G3789797				
		ACTIN BINDING PROTEIN MAYVEN. Length				
		= 593				
		heparin-binding fibroblast growth factor receptor				
		2 [Rattus norvegicus] >sp Q63241 Q63241				
		HEPARIN-BINDING FIBROBLAST GROWTH				
		FACTOR RECEPTOR 2 (FRAGMENT). {SUB				
		1-330} Length = 331				
		15 KDA SELENOPROTEIN. Length = 162				
			sp O60613 O60613	3	374	IACGIY70

771	844408	(AF001437) dihydrolipoamide dehydrogenase-binding protein [Homo sapiens] Length = 501	gi 2316040	1358	1651	100	100	HTNAD87
772	844508	pre-pro polypeptide (AA -25 to 451) [Homo sapiens] >pir S09489 S09489 carboxypeptidase H (EC 3.4.17.10) precursor - human		1	300			IIADGG65
773	844867	>sp P16870 CBPH_HUMAN		174	371			IIMVBJ82
774	845000	CARBOXYPEPTIDASE H PRECURSOR (EC 3.4.17.10) (CPH) (CARBOXYPEPTIDASE E) (CPE) (ENKEPHALIN CONVERTASE)		1	321			HE9DB89
775	845281	(PROHORMON)	gi 29667	3	1475	100	100	III:GAE94
776	845288	(AF023268) propin I [Homo sapiens] Length = 347	gi 2564915	571	1107	75	76	ITLDM37
777	845750	selenium-binding protein [Homo sapiens] >pir G01872 G01872 selenium-binding protein - human >sp Q13228 Q13228 SELENIUM-BINDING PROTEIN. Length = 472	gi 1374792	3	1499	95	96	IIIE9DH28
778	845809	SNAP23A protein [Homo sapiens] >gnl PID e1331767 (AJ011915) synaptosome associated protein of 23 kilodaltons, isoform A [Homo sapiens] >pir JC5296 JC5296 vesicle-membrane fusion protein SNAP-23A - human >sp O00161 O00161 VESICLE-MEMBRANE FUSION PROTEIN SN	gnl PID e290695	134	772	100	100	IIKGE41
779	846077			182	487			HCNCN11
780	HPFCH77R			21	80			HPFCH77
781	HPRT105R			2	151			HPRT105
782	HMSKI93R			25	192			HMSKI93

80

783	IIKAAAC88R (AB003103) 26S proteasome subunit p55 [Homo sapiens] >sp O00232 O00232 PROTEASOME SUBUNIT P55. Length = 456	gn PID d1020530	1	333	85	88	IIKAAAC88
784	HPDED94R (AF001212) 26S proteasome subunit 9 [Homo sapiens] >sp O00495 O00495 26S PROTEASOME SUBUNIT 9. Length = 422	gi 2150046	1	225	98	98	HPDED94
785	IIDTGH11R (AF009674) axin [Homo sapiens] >sp O15169 O15169 AXIN (FRAGMENT). Length = 900	gi 2252820	1	189	96	96	IIDTGH11
786	IITEJR60R (AF022184) EZF [Homo sapiens] >sp O43474 EZF_HUMAN EPITHELIAL ZINC-FINGER PROTEIN EZF. Length = 470	gi 2897954	2	511	77	77	IITEJR60
787	IIAGGY86R (AF029786) GBAS [Homo sapiens] >sp O75323 O75323 GBAS. Length = 286	gi 3403167	2	295	97	98	IIAGGY86
788	IIPIAU47R (AF031647) 1AB1-containing signalosome subunit 3 [Homo sapiens] >sp O43191 O43191 SIGNALOSOME SUBUNIT 3. Length = 403	gi 2688989	3	377	89	91	IIPIAU47
789	HCGAD89R (AF074935) beta-tubulin [Cryptosporidium parvum] >gi 3328337 (AF074936) beta-tubulin [Cryptosporidium parvum] >sp O77467 O77467 BETA-TUBULIN (FRAGMENT). Length = 57	gi 3328335	226	390	86	89	HCGAD89
790	IIAPOD39R (AF089866) keratin 19 [Rattus norvegicus] >sp G3766220 G3766220 KERATIN 19 (FRAGMENT). Length = 123	gi 3766220	3	386	88	93	IIAPOD39
791	IIOGAA68R 5' half of the product is homologues to Bacillus subtilis SAICAR synthetase, 3' half corresponds to the catalytic subunit of AIR carboxylase [Homo sapiens] >pir S14147 S14147 multifunctional purine biosynthesis protein - human Length = 425	gi 28384	1	468	95	97	IIOGAA68

81

792	HCLBO46R	Actin [Drosophila melanogaster] >pir S14851 S14851 actin - fruit fly (Drosophila melanogaster) >sp Q24228 Q24228 ACTIN. Length = 100	gi 7550	1	303	94	95	HCLBO46
793	IIDRAA14R	ADP,ATP carrier protein T2 - human >sp P12236 ADT3_HUMAN ADP,ATP CARRIER PROTEIN, LIVER ISOFORM T2 (ADP/ATP TRANSLOCASE 3) (ADENINE NUCLEOTIDE TRANSLOCATOR 3) (ANT 3). Length = 298	pir S03894 S03894	2	304	80	92	IIDRAA14
794	HSLCA48R	alpha-1 (III) collagen [Homo sapiens] Length = 1078	gi 930045	2	457	70	75	HSLCA48
795	IIMEAC81R	alpha-subunit of G-protein, type G-alpha-i-1 [Xenopus laevis] >pir S11045 RGXLI1 GTP- binding regulatory protein Gi alpha-1 chain (adenylate cyclase-inhibiting) - African clawed frog >sp P27044 GBI1_XENLA GUANINE NUCLEOTIDE-BINDING PROTEIN G(I), ALPHA-1 SU	gi 64708	99	176	92	92	IIMEAC81
796	IIMQDF20R	beta-1,2-N-acetylglucosaminyltransferase II [Homo sapiens] >pir S66256 S66256 alpha-1,6- mannosyl-glycoprotein beta-1, 2-N- acetylglucosaminyltransferase (EC 2.4.1.143) - human >sp Q10469 GNT2_HUMAN ALPHA- 1,6-MANNOSYL-GLYCOPROTEIN BETA-1,2- N- ACETYLGLUCOSAM	gi 902745	3	287	85	85	IIMQDF20
797	IICHOH06R			12	242			IICHOH06
798	HDQMC20R			3	167			HDQMC20
799	IIMKCW11R			2	112			IIMKCW11

800	HILDRN91R	C4b-binding protein alpha chain [Homo sapiens] >gil190502 C4b-binding protein alpha chain [Homo sapiens] >pirA33568 NBHUC4 C4b- binding protein alpha chain precursor - human >spP04003 C4BP_HUMAN C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR (PROLINE-RICH PRO	gil190500	2	331	99	100	HILDRN91
801	HICHR17R	cathepsin D [Homo sapiens] >gil29678 precursor polypeptide (AA -20 to 392) [Homo sapiens] >gil181180 preprocathepsin D [Homo sapiens] >pirA2577 KHHUD cathepsin D (EC 3.4.23.5) precursor - human >spP07339 CATD_HUMAN CATHEPSIN D PRECURSOR (EC 3.4.23.5).	gil179948	3	149	92	92	HICHR17
802	HMKCH15R	Cbf5p homolog [Homo sapiens] Length = 514	gil2737894	131	400	81	81	HMKCH15
803	HIE6GO78R	clathrin light-chain A [Homo sapiens] Length = 218	gil307118	155	502	80	83	HIE6GO78
804	HISL1F156R	complement component C3 [Homo sapiens] >pirA94065 C3HU complement C3 precursor - human >spP01024 CO3_HUMAN COMPLEMENT C3 PRECURSOR [CONTAINS: C3A ANAPHYLATOXIN]. >gil181130 complement component C3 [Homo sapiens] {SUB 1-24} Length = 1663	gil179665	48	422	80	82	HISL1F156
805	HISYBY17R	cyclin G [Homo sapiens] >gil1236233 cyclin G1 [Homo sapiens] >gil1236913 cyclin G1 [Homo sapiens] >pirG02401 G02401 cyclin G1 - human >spP51959 CG2G_HUMAN G2/MITOTIC- SPECIFIC CYCLIN G1. >gnlPID1013694 cyclin G [Homo sapiens] {SUB 1-279} >gil1486361 c	gnlPID1012016	79	300	100	100	HISYBY17

806	IIPJCS07R	cytochrome oxidase I [Apteryx australis] >sp O03515 COX1_APTAU CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT). Length = 337	gij2198683	113	226	83	92	IIPJCS07
807	IIFADV82R	cytochrome oxidase III [Homo sapiens] >pir A00482 OTHU3 cytochrome-c oxidase (EC 1.9.3.1) chain III - human mitochondrion (SGC1) >sp P00414 COX3_HUMAN CYTOCHROME C OXIDASE POLYPEPTIDE III (EC 1.9.3.1). >gij2245564 (AF004341) cytochrome c oxidase subunit I	gij13010	1	105	81	83	IIFADV82
808	IIFKFFH08R	DNA polymerase delta small subunit [Homo sapiens] >pir J38950 J38950 DNA-directed DNA polymerase (EC 2.7.7.7) delta regulatory chain - human >sp P49005 DPD_HUMAN DNA POLYMERASE DELTA SMALL SUBUNIT (EC 2.7.7.7). Length = 469	gij1008458	2	550	97	98	IIFKFFH08
809	HMC DK47R	electron transport flavoprotein [Homo sapiens] >pir A31998 A31998 electron transfer flavoprotein alpha chain precursor - human >sp P13804 ETFA_HUMAN ELECTRON TRANSFER FLAVOPROTEIN ALPHA- SUBUNIT PRECURSOR (ALPHA-ETF). >gnl PID1e1331769 (AJ224002) electron elongation factor 2 [Homo sapiens] >gij31108 human elongation factor 2 [Homo sapiens] >pir S18294 EFHU2 translation elongation factor eEF-2 - human >sp P13639 EF2_HUMAN ELONGATION FACTOR 2 (EF-2). >gij181969 elongation factor 2 [Homo sapiens] {SUB 501- 858	gij182251	3	320	100	100	HMC DK47
810	IIPB127R		gij31106	23	319	98	98	IIPB127

84

811	IIISKJG37R	elongation factor 2 [Homo sapiens] >gil31108 human elongation factor 2 [Homo sapiens] >pir S18294 EFHU2 translation elongation factor eEF-2 - human >sp P13639 EF2_HUMAN ELONGATION FACTOR 2 (EF-2). >gil181969 elongation factor 2 [Homo sapiens] {SUB 501- 858	gil31106	1	372	100	100	100	HSKJG37
812	II2LAZ24R	elongation factor-1-beta [Homo sapiens] >gil31135 elongation factor 1-beta [Homo sapiens] >pir S25432 S25432 translation elongation factor eEF-1 beta chain - human >sp P24534 EF1B_HUMAN ELONGATION FACTOR 1-BETA (EF-1-BETA). {SUB 2-225} Length = 225	gil31100	23	562	100	100	100	II2LAZ24
813	II2LAC50R	enhancer protein [Homo sapiens] >pir S4533 S4533 enhancer protein - human Length = 199	gil440306	38	415	100	100	100	II2LAC50
814	IIPEAE15R	GLANDULAR KALLIKREIN-1. Length = 223	sp Q15946 Q15946	51	236	80	80	80	IIPEAE15
815	IIPIAA24R	GTP-binding protein Ran/TC4 - mouse (fragment) Length = 70	pir JH0654 JH0654	382	507	91	91	91	IIPIAA24
816	II2LAS11R	guanylate cyclase (EC 4.6.1.2) - bovine (fragment) >gil407777 guanylate cyclase [Bos taurus] {SUB 2-498} Length = 498	pir S48119 S48119	28	549	100	100	100	II2LAS11
817	IIIIERW66R	HMG1 protein (AA 1 - 215) [Bos taurus] >pir S01947 S01947 nonhistone chromosomal protein HMG-1 - bovine >sp P10103 HMG1_BOVIN HIGH MOBILITY GROUP PROTEIN HMG1 (HMG-1). {SUB 2- 215} Length = 215	gil417	3	386	83	83	83	IIIIERW66

818	HADMC73R	hMn-superoxiddismutase [unidentified] >gi 491292 hMn-superoxiddismutase [unidentified] >gnl PID e93456 Mn- superoxiddismutase [Homo sapiens] {SUB 23- 199} Length = 199	gi 491290	2	94	96	100	HADMC73
819	H6EEU22R	hormone receptor hERR1 (AA 1-521) [Homo sapiens] >pir A29345 A29345 steroid hormone receptor ERR1 precursor - human >sp P11474 ERR1_HUMAN STEROID HORMONE RECEPTOR ERR1 (ESTROGEN- RELATED RECEPTOR, ALPHA) (ESTROGEN RECEPTOR-LIKE 1). Length = 521	gi 36609	34	225	100	100	H6EEU22
820	IIDTDX66R	HPIHs-gamma [Homo sapiens] >sp Q13185 HPIG_HUMAN HETEROCHROMATIN PROTEIN 1 HOMOLOG GAMMA (HPI GAMMA) (MODIFIER 2 PROTEIN). >sp G1773227 G1773227 HPIHS-GAMMA. Length = 173	gi 1773227	132	449	82	84	IIDTDX66
821	ILPBB39R	human metallothionein-1e [Homo sapiens] >pir A22634 SMHUIE metallothionein 1E - human >sp P04732 MT1E_HUMAN METALLOTHIONEIN-1E (MT-1E). >bbs 144157 metallothionein MT-1e isoform, metallothionein-1e [human, monocytes, Peptide Partial, 31 aa] [Homo sapiens]	gi 386865	40	246	100	100	ILPBB39
822	HOELG04R	hypothetical 18K protein (rRNA) - goldfish mitochondrion (SGC1) Length = 166	pir JC1348 JC1348	293	415	65	68	HOELG04

86

823	IIKABU38R	initiation factor 4B [Homo sapiens] >pir S12566 S12566 translation initiation factor eIF-4B - human >sp P23588 IF4B_HUMAN EUKARYOTIC TRANSLATION INITIATION FACTOR 4B (EIF-4B). Length = 611	gi 288100	2	463	92	92	IIKABU38
824	IIBGOI32R	keratin 18 [Homo sapiens] >gi 307081 keratin 18 precursor [Homo sapiens] >gi 34037 cytokeratin 18 [Homo sapiens] >pir S05481 S05481 keratin 18, type I, cytoskeletal - human >sp P05783 K1CR_HUMAN KERATIN, TYPE I CYTOSKELETAL 18 (CYTOKERATIN 18) (K18) (CK 1	gi 386844	1	240	66	67	IIBGOI32
825	IIATAI03R	KIAA0106 [Homo sapiens] >sp P30041 AOP2_HUMAN ANTIOXIDANT PROTEIN 2 (EC 1.1.1.7) (24 KD PROTEIN) (LIVER 2D PAGE SPOT 40) (RED BLOOD CELLS PAGE SPOT 12). {SUB 2-224} Length = 224	gnl PID d1004007	3	194	90	93	IIATAI03
826	ICEDE25R	KIAA0106 [Homo sapiens] >sp P30041 AOP2_HUMAN ANTIOXIDANT PROTEIN 2 (EC 1.1.1.7) (24 KD PROTEIN) (LIVER 2D PAGE SPOT 40) (RED BLOOD CELLS PAGE SPOT 12). {SUB 2-224} Length = 224	gnl PID d1004007	2	283	100	100	ICEDE25
827	HKDBF62R	metallothionein-IG [Homo sapiens] >pir A29236 SMHU1G metallothionein 1G - human >sp P13640 MT1G_HUMAN METALLOTHIONEIN-IG (MT-IG). >bls 14160 metallothionein MT-1g isoform, metallothionein-1g [human, monocytes, Peptide Partial, 31 aa] [Homo sapiens] {SUB	gi 188713	170	322	95	95	HKDBF62

828	IINTSX94R	mitochondrial matrix protein [Homo sapiens] >pir A32800 A32800 chaperonin GroEL precursor - human >sp P10809 P60_HUMAN MITOCHONDRIAL MATRIX PROTEIN P1 PRECURSOR (P60 LYMPHOCYTE PROTEIN) (60 KD CHAPERONIN) (HEAT SHOCK PROTEIN 60) (HSP-60) (PROTEIN CPN60) (	3	431	97	100	IINTSX94
829	IIRGBR08R	mitochondrial matrix protein [Homo sapiens] >pir A32800 A32800 chaperonin GroEL precursor - human >sp P10809 P60_HUMAN MITOCHONDRIAL MATRIX PROTEIN P1 PRECURSOR (P60 LYMPHOCYTE PROTEIN) (60 KD CHAPERONIN) (HEAT SHOCK PROTEIN 60) (HSP-60) (PROTEIN CPN60) (	1	504	94	94	IIRGBR08
830	I12LAO77R	MSS1 protein [Homo sapiens] >pir S24353 S24353 proteasome 26S subunit MSS1 - human >sp G385267 G385267 26 S PROTEASE SUBUNIT 7, MSS1=MODULATOR OF HIV TAT- MEDIATED TRANSACTIVATION. {SUB 2- 23} Length = 433	137	580	91	91	I12LAO77R
831	IINTRW15R	NAD+ ADP-ribosyltransferase [Homo sapiens] >pir A29725 A29725 NAD+ ADP- ribosyltransferase (EC 2.4.2.30), nuclear - human >sp P09874 PPOL_HUMAN POLY [ADP- RIBOSE] POLYMERASE (EC 2.4.2.30) (PARP) (ADPRT) (NAD(+)) ADP- RIBOSYLTRANSFERASE (POLY[ADP- RIBOSE] SYN	163	297	90	96	IINTRW15

832	HORBH08R	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 51K chain precursor - human (fragment) >sp P4982 INUBM_HUMAN NADH-UBIQUINONE OXIDOREDUCTASE 51 KD SUBUNIT PRECURSOR (EC 1.6.5.3) (EC 1.6.99.3) (COMPLEX I-51KD) (CI-51KD) (FRAGMENT). >bbs 142159 NADH:ubiquinone nonstructural protein P125-2 [pestivirus type 1] >sp O57114 O57114 NONSTRUCTURAL PROTEIN P125-2 (FRAGMENT). Length = 239	pir A44362 A44362	186	428	83	87	HORBH08
833	HULBL38R	p60 [Homo sapiens] >sp Q13446 Q13446 EB13-ASSOCIATED PROTEIN P60. >gil 3283216 (AF060494) ubiquitin binding protein p62 [Homo sapiens] {SUB 1-72} Length = 440	gil 2707597	3	437	95	97	HULBL38
834	INTBK49R	Phalaenopsis sp. 'hybrid SM9108' actin [Phalaenopsis sp. 'hybrid SM9108'] >sp Q40981 Q40981 ACTIN (FRAGMENT). Length = 208	gil 145799	3	368	100	100	INTBK49
835	IIBAFS48R	PIPpin protein [Rattus norvegicus] >pir JC4588 JC4588 RNA-binding protein PIPpin - rat >sp Q63430 Q63430 PIPPIN PROTEIN. Length = 154	gil 602958	2	316	91	92	IIBAFS48
836	HHGAL60R	prepro-alpha-1 collagen [Homo sapiens] >sp Q15201 Q15201 PREPRO-ALPHA-1 COLLAGEN PRECURSOR (FRAGMENT). Length = 181	gil 1050754	2	319	66	81	HHGAL60
837	HOHBU75R	progesterone-induced protein [Oryctolagus cuniculus] >pir A26998 A26998 progesterone-induced protein, endometrial - rabbit Length = 370	gil 35658	104	373	71	72	HOHBU75
838	IIHEFZ79R		gil 165009	293	484	73	77	IIHEFZ79

839	IISLBA61R	proteasome subunit C5 [Homo sapiens] >gnl PID e1334433 (AL031259) C5 (proteasome subunit HC5) [Homo sapiens] >pir S15973 SNHUC5 multicatalytic endopeptidase complex (EC 3.4.99.46) chain C5 - human >sp P20618 PRC5_HUMAN PROTEASOME COMPONENT C5 (EC 3.4.99.4 pui. ORF [Homo sapiens] >pir I38022 I38022 hypothetical protein - human >sp Q29976 Q29976 MAHLAVU HEPATOCELLULAR CARCINOMA HHC(M) DNA. Length = 196 ren(exclusion;96) [Bacteriophage lambda] >pir F43010 ZRRBPL ren protein - phage lambda Length = 96 ribosomal protein small subunit [Homo sapiens] Length = 264 ribosomal phosphoprotein P1 (AA 1-114) [Rattus rattus] >pir S08022 RSRT12 acidic ribosomal protein P1 - rat Length = 114 ribosomal protein L10 [Homo sapiens] >sp D102677 D102677 RIBOSOMAL PROTEIN L15 (FRAGMENT). (SUB 16-57) Length = 205 ribosomal protein L15 gene product [Rattus norvegicus] >pir JC2369 JC2369 ribosomal protein L15 - rat Length = 204	gnl PID d1001116	45	224	96	96	IISLBA61
840	HPEAE18R		gj 288145	55	234	57	67	HPEAE18
841	HNGFO65R		gj 215152	3	203	48	59	HNGFO65
842	HKAKR61R		gj 306553	3	458	91	91	HKAKR61
843	I2LAP11R		gj 57710	169	549	100	100	I2LAP11
844	H2CBD90R		gj 414587	199	501	95	95	H2CBD90
845	H2LAD40R		gj 515865	156	524	100	100	H2LAD40

90

846	HCYBK51R	ribosomal protein L37 [Homo sapiens] >bbs 172744 ribosomal protein L37 {C2-C2 zinc-finger-like} [human, HeLa cells, Peptide, 97 aa] [Homo sapiens] >gnl PID d1005426 ribosomal protein L37 [Homo sapiens] >gij 57121 ribosomal protein L37 [Rattus norvegicus] >	gi 292441	2	412	97	98	HCYBK51
847	I12MBC73R	ribosomal protein L37a [Homo sapiens] >gij 36134 ribosomal protein L37a [Homo sapiens] >gij 57123 ribosomal protein L37a (AA 1 - 92) [Rattus rattus] >gij 312414 ribosomal protein L37a [Mus musculus] >pir S05014 R5RT37 ribosomal protein L37a - rat >pir S42109	gi 292439	2	385	100	100	I12MBC73
848	I12MBU27R	ribosomal protein L37a [Homo sapiens] >gij 36134 ribosomal protein L37a [Homo sapiens] >gij 57123 ribosomal protein L37a (AA 1 - 92) [Rattus rattus] >gij 312414 ribosomal protein L37a [Mus musculus] >pir S05014 R5RT37 ribosomal protein L37a - rat >pir S42109	gi 292439	2	286	100	100	I12MBU27
849	I12SAH53R	ribosomal protein L37a [Homo sapiens] >gij 36134 ribosomal protein L37a [Homo sapiens] >gij 57123 ribosomal protein L37a (AA 1 - 92) [Rattus rattus] >gij 312414 ribosomal protein L37a [Mus musculus] >pir S05014 R5RT37 ribosomal protein L37a - rat >pir S42109	gi 292439	3	341	97	97	I12SAH53
850	HAIDF69R	ribosomal protein L7a [Fugu rubripes] Length = 266	gnl PID e1248480	179	250	93	100	HAIDF69

851	HDBAA15R	ribosomal protein L8 [Homo sapiens] >gi 57704 ribosomal protein L8 [Rattus rattus] >gi 527178 ribosomal protein L8 [Mus musculus] >pir JU0177 R5RTL8 ribosomal protein L8, cytosolic - rat >pir JN0923 JN0923 ribosomal protein L8, cytosolic - human >gi 3851	gi 433899	220	429	85	88	HDBAA15
852	HIDTHW54R	ribosomal protein S12 (AA 1 - 132) [Mus musculus] >pir S13074 R3R112 ribosomal protein S12 - rat >pir S05492 R3MS12 ribosomal protein S12 - mouse >gi 206741 ribosomal protein S12 [Rattus norvegicus] {SUB 1-130} Length = 132	gi 54006	3	332	89	89	HIDTHW54
853	HTWJC11R	ribosomal protein S13 [Homo sapiens] >gi 488417 ribosomal protein S13 [Homo sapiens] >gnl PID d1014222 ribosomal protein S13 [Homo sapiens] >gi 57730 ribosomal protein S13 [Rattus rattus] >pir S34109 S34109 ribosomal protein S13, cytosolic - human >pir A3	gi 307391	1	276	97	97	HTWJC11
854	HKAEC40R	ribosomal protein S24 [Homo sapiens] >gi 517222 ribosomal protein S24 [Homo sapiens] >gi 49652 ribosomal protein S19 (AA 1 - 133) [Mesocricetus auratus] >gi 57858 ribosomal protein S24 [Rattus norvegicus] >gi 57722 ribosomal protein S24 (AA 1-133) [Rattus	gi 337506	93	407	83	84	HKAEC40
855	ICFNM70R	ribosomal protein S4X isoform [Homo sapiens] >gi 2791861 (AF041428) ribosomal protein s4 X isoform [Homo sapiens] >gi 200864 ribosomal protein S4 [Mus musculus] >gi 57135 ribosomal protein S4 (AA 1 - 263) [Rattus rattus] >gnl PID d1002335 ribosomal protei	gi 337510	3	278	96	97	ICFNM70

92

856	HKBAB93R	ribosomal protein S8 [Homo sapiens] >gi 57139 ribosomal protein S8 (AA 1-208) [Rattus norvegicus] >gi 313298 ribosomal protein S8 [Mus musculus] >pir S01609 R3RT8 ribosomal protein S8 - rat >pir S42110 S42110 ribosomal protein S8 - mouse >pir S25022 S2502	gi 36150	2	391	87	90	HKBAB93
857	IILHEJ79R	RNA polymerase II subunit hRPB17 [Homo sapiens] >pir S55370 S55370 RNA polymerase II chain hRPB17 - human Length = 150	gi 854177	129	446	83	86	HLHE179
858	IIBGOI24R	S19 ribosomal protein [Homo sapiens] >pir S2692 S2692 ribosomal protein S19, cytosolic - human Length = 145	gi 337733	2	421	99	100	HBG0124
859	IINDAD16R	secretory protein [Homo sapiens] >gi 940946 intestinal trefoil factor [Homo sapiens] >pir A48284 A48284 intestinal trefoil factor 3 precursor - human >sp Q07654 TF_HUMAN INTESTINAL TREFOIL FACTOR PRECURSOR (HP1.B). Length = 80	gi 402483	3	380	71	78	IINDAD16
860	HIMAE94R	serine/threonine protein kinase [Homo sapiens] >gnl PID e1154172 (AJ000512) serine/threonine protein kinase [Homo sapiens] Length = 431	gnl PID e293330	3	422	95	95	HIMAE94
861	IIMWEA08R	signal recognition particle subunit 9 [Homo sapiens] >pir A57292 A57292 signal recognition particle protein SRP9 - human Length = 86	gi 897851	119	394	90	93	IIMWEA08
862	I16BSO48R	similar to Drosophila photoreceptor cell-specific protein, calphotin. [Homo sapiens] >sp Q14676 Q14676 KIAA0170 PROTEIN. Length = 2089	gnl PID d1012153	1	528	95	95	I16BSO48

863	HIRACC09	smooth muscle protein [Homo sapiens] >pir J50774 J50774 smooth muscle protein SM22 - human Length = 201	gij 177175	1	117	100	100	HIRACC09
864	HOEEC67R	smooth muscle protein SM22 homolog - mouse Length = 201	pir A60598 A60598	105	230	100	100	HOEEC67
865	HPFEA40R	t-complex polypeptide 1 (AA 1-556) [Homo sapiens] Length = 556	gij 36796	3	497	98	99	HPFEA40
866	HODAV31R	tissue inhibitor of metalloproteinases [Homo sapiens] Length = 166	gnl P1D1d I002390	1	273	64	67	HODAV31
867	HIHEC189R	transaldolase [Homo sapiens] >gij 2612879 (AF010400) transaldolase-related protein [Homo sapiens] >sp O00751 O00751 TRANSALDOLASE (EC 2.2.1.2). >gij 480787 transaldolase [Homo sapiens] {SUB 302-337} Length = 337	gij 2073541	3	371	99	99	HIHEC189
868	HSDFV03R	translocase [Bos taurus] >pir B43646 B43646 ADP, ATP carrier protein T2 - bovine >sp P32007 ADT3_BOVIN ADP, ATP CARRIER PROTEIN, ISOFORM T2 (ADP/ATP TRANSLOCASE 3) (ADENINE NUCLEOTIDE TRANSLOCATOR 3) (ANT 3). Length = 298	gij 529417	20	412	92	96	HSDFV03
869	HTXPN01R	triose-phosphate isomerase [Pan troglodytes] >gij 37247 triosephosphate isomerase [Homo sapiens] >gij 200507 triosephosphate isomerase [Homo sapiens] >gij 339841 triosephosphate isomerase (EC 5.3.1.1) [Homo sapiens] >pir S29743 ISHUT triose-phosphate isomer	gij 176960	3	281	98	98	HTXPN01
870	HHPSA49R	tuberin [Homo sapiens] Length = 1784	gij 450352	2	451	69	69	HHPSA49
871	HI2LAT88R	type II mesothelial keratin K7 [Homo sapiens] >sp Q92676 Q92676 MESOTHELIAL KERATIN K7 (TYPE II) (FRAGMENT). Length = 489	gij 386851	1	567	91	91	HI2LAT88

95

872	H6EAD58R	49	174	IIEAD58
873	HACBH95R	2	364	HACBH95
874	IACBY16R	1	84	IACBY16
875	HAGCI33R	2	238	HAGCI33
876	HAHAD34R	61	123	HAHAD34
877	IAJAN69R	67	294	IAJAN69
878	IALSG52R	41	268	IALSG52
879	HAPPR17R	180	311	HAPPR17
880	HAQCG78R	3	110	HAQCG78
881	IAUBY86R	23	118	IAUBY86
882	HAVAA34R	1	117	HAVAA34
883	HBAFK20R	2	355	HBAFK20
884	HGBE20R	31	315	HGBE20
885	HBJBR66R	2	52	HBJBR66
886	HBJMU59R	2	208	HBJMU59
887	HBKDK63R	147	647	HBKDK63
888	HBMVT43R	2	70	HBMVT43
889	HCDAM59R	21	125	HCDAM59
890	HCFLN25R	3	224	HCFLN25
891	HCQAW59R	1	129	HCQAW59
892	IIDPMA46R	223	420	IIDPMA46
893	IIDTAQ26R	177	296	IIDTAQ26
894	HDATAT40R	1	213	HDATAT40
895	IDTLD39R	323	496	IDTLD39
896	HE2PO63R	39	278	HE2PO63
897	HELCV09R	1	72	HELCV09
898	HELHK95R	3	383	HELHK95
899	HEMGL70R	2	172	HEMGL70
900	HEIIB72R	2	100	HEIIB72
901	HFFAS19R	2	256	HFFAS19
902	HF1YH65R	68	259	HF1YH65

903	HFXAF89R	143	361	HFXAF89
904	HHEPR03R	89	307	HHEPR03
905	HHGAQ80R	2	202	HHGAQ80
906	HHSEF82R	170	304	HHSEF82
907	HKBA63R	239	469	HKBA63
908	HKIXO47R	2	94	HKIXO47
909	IILDNF70R	3	176	IILDNF70
910	IIIQFO33R	62	268	IIIQFO33
911	HLWBC80R	46	543	HLWBC80
912	IILYAV50R	3	224	IILYAV50
913	HMEKY67R	3	302	HMEKY67
914	IIMTBN58R	3	377	IIMTBN58
915	HNGAZ91R	22	276	HNGAZ91
916	HNTAC06R	2	133	HNTAC06
917	HOGAF41R	1	228	HOGAF41
918	HOUDQ92R	75	323	HOUDQ92
919	HPEAD91R	60	233	HPEAD91
920	HPIAF72R	128	310	HPIAF72
921	IPIAU01R	122	334	IPIAU01
922	IPIAU73R	99	275	IPIAU73
923	HPIAW19R	102	350	HPIAW19
924	IPIAZ19R	238	348	IPIAZ19
925	HPIBA31R	245	367	HPIBA31
926	IPIBS06R	84	182	IPIBS06
927	IPICB65R	2	430	IPICB65
928	IPIJBF22R	220	330	IPIJBF22
929	IPIJBZ81R	214	384	IPIJBZ81
930	HRACF81R	1	189	HRACF81
931	HIRACT28R	110	319	HIRACT28
932	IISBAP03R	123	263	IISBAP03
933	IISDJK57R	234	458	IISDJK57

934	HSIFY54R	1	321	HSIFY54
935	HSLDJ92R	24	275	HSLDJ92
936	HSLJI47R	185	379	HSLJI47
937	HTSGE55R	36	209	HTSGE55
938	HUFAT72R	276	410	HUFAT72
939	HULAI70R	176	337	HULAI70
940	HTGFW12R	3	233	HTGFW12

gnl|PID|d1008092

yeast mismatch repair gene PMS1 homologue
 [Homo sapiens] >gnl|PID|d1008050 homologue
 of yeast PMS1 [Homo sapiens]
 >sp|Q16530|Q16530 PMS3 MRNA (YEAST
 MISMATCH REPAIR GENE PMS1
 HOMOLOGUE), PARTIAL CDS (C-
 TERMINAL REGION) (FRAGMENT). Length =
 256

The first column of Table 1 shows the "SEQ ID NO:" for each of the 940 prostate cancer antigen polynucleotide sequences of the invention.

The second column in Table 1, provides a unique "Sequence/Contig ID" identification for each prostate and/or prostate cancer associated sequence. The third column in Table 1, "Gene Name," provides a putative identification of the gene based on the sequence similarity of its translation product to an amino acid sequence found in a publicly accessible gene database, such as GenBank (NCBI). The great majority of the cDNA sequences reported in Table 1 are unrelated to any sequences previously described in the literature. The fourth column, in Table 1, "Overlap," provides the database accession no. for the database sequence having similarity. The fifth and sixth columns in Table 1 provide the location (nucleotide position nos. within the contig), "Start" and "End", in the polynucleotide sequence "SEQ ID NO:X" that delineate the preferred ORF shown in the sequence listing as SEQ ID NO:Y. In one embodiment, the invention provides a protein comprising, or alternatively consisting of, a polypeptide encoded by the portion of SEQ ID NO:X delineated by the nucleotide position nos. "Start" and "End". Also provided are polynucleotides encoding such proteins and the complementary strand thereto. The seventh and eighth columns provide the "% Identity" (percent identity) and "% Similarity" (percent similarity) observed between the aligned sequence segments of the translation product of SEQ ID NO:X and the database sequence.

The ninth column of Table 1 provides a unique "Clone ID" for a clone related to each contig sequence. This clone ID references the cDNA clone which contains at least the 5' most sequence of the assembled contig and at least a portion of SEQ ID NO:X was determined by directly sequencing the referenced clone. The reference clone may have more sequence than described in the sequence listing or the clone may have less. In the vast majority of cases, however, the clone is believed to encode a full-length polypeptide. In the case where a clone is not full-length, a full-length cDNA can be obtained by methods described elsewhere herein.

Table 3 indicates public ESTs, of which at least one, two, three, four, five, ten, or more of any one or more of these public ESTs are optionally excluded from the invention.

SEQ ID NO:X (where X may be any of the polynucleotide sequences disclosed in the sequence listing as SEQ ID NO:1 through SEQ ID NO:940) and the translated SEQ ID NO:Y (where Y may be any of the polypeptide sequences disclosed in the sequence listing as SEQ ID NO:941 through SEQ ID NO:1880) are sufficiently accurate and otherwise suitable for a

variety of uses well known in the art and described further below. For instance, SEQ ID NO:X has uses including, but not limited to, in designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the related cDNA clone contained in a library deposited with the ATCC. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling immediate applications in chromosome mapping, linkage analysis, tissue identification and/or typing, and a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y have uses that include, but are not limited to, generating antibodies which bind specifically to the prostate cancer antigen polypeptides, or fragments thereof, and/or to the prostate cancer antigen polypeptides encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X, the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing the related cDNA clone (deposited with the ATCC, as set forth in Table 1). The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. Further, techniques known in the art can be used to verify the nucleotide sequences of SEQ ID NO:X.

The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to vectors or plasmids which include such DNA sequences, as well as the use of the DNA sequences. The material deposited with the ATCC on:

5 **Table 2**

ATCC Deposits	Deposit Date	ATCC Designation Number
LP01, LP02, LP03, LP04, LP05, LP06, LP07, LP08, LP09, LP10, LP11.	May-20-97	209059, 209060, 209061, 209062, 209063, 209064, 209065, 209066, 209067, 209068, 209069
LP12	Jan-12-98	209579
LP13	Jan-12-98	209578
LP14	Jul-16-98	203067
LP15	Jul-16-98	203068
LP16	Feb-1-99	203609
LP17	Feb-1-99	203610
LP20	Nov-17-98	203485
LP21	Jun-18-99	PTA-252
LP22	Jun-18-99	PTA-253
LP23	Dec-22-99	PTA-1081

each is a mixture of cDNA clones derived from a variety of human tissue and cloned in either a plasmid vector or a phage vector, as shown in Table 5. These deposits are referred to as "the deposits" herein. The tissues from which the clones were derived are listed in Table 5, and the vector in which the cDNA is contained is also indicated in Table 5. The deposited material includes the cDNA clones which were partially sequenced and are related to the SEQ ID NO:X described in Table 1 (column 9). Thus, a clone which is isolatable from the ATCC Deposits by use of a sequence listed as SEQ ID NO:X may include the entire coding region of a human gene or in other cases such clone may include a substantial portion of the coding region of a human gene. Although the sequence listing lists only a portion of the DNA sequence in a clone included in the ATCC Deposits, it is well within the ability of one

ATCC Deposits by use of a sequence (or portion thereof) listed in Table 1 by procedures hereinafter further described. and others apparent to those skilled in the art.

Also provided in Table 5 is the name of the vector which contains the cDNA clone. Each vector is routinely used in the art. The following additional information is provided for
5 convenience.

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., *Nucleic Acids Res.* 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., *Nucleic Acids Res.* 17:9494 (1989)) and pBK (Alting-
10 Mees, M. A. et al., *Strategies* 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Phagemid pBS may be excised from the Lambda Zap and Uni-Zap XR vectors, and phagemid pBK may be excised from the Zap Express vector. Both phagemids may be transformed into *E. coli* strain
15 XL-1 Blue, also available from Stratagene.

Vectors pSport1, pCMVSPORT 1.0, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, also available from Life Technologies. See, for instance, Gruber, C. E., et al., *Focus*
20 15:59 (1993). Vector lafmid BA (Bento Soares, Columbia University, New York, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue. Vector pCR[®]2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies. See, for instance, Clark, J. M., *Nuc. Acids Res.*
25 16:9677-9686 (1988) and Mead, D. et al., *Bio/Technology* 9: (1991).

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, and/or the cDNA contained in a deposited cDNA clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include, but are not limited to, preparing probes or primers from the
30 disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are allelic variants, orthologs, and/or species homologs. Procedures known in the art can be used to obtain full-length genes, allelic variants, splice variants, full-length coding portions, orthologs, and/or species homologs of genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, and/or the cDNA contained in the related cDNA clone in the deposit, using information from the sequences disclosed herein or the clones deposited with the ATCC. For example, allelic variants and/or species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for allelic variants and/or the desired homologue.

The present invention provides a polynucleotide comprising, or alternatively consisting of, the nucleic acid sequence of SEQ ID NO:X, and/or the related cDNA clone (See, e.g., columns 1 and 9 of Table 1). The present invention also provides a polypeptide comprising, or alternatively, consisting of, the polypeptide sequence of SEQ ID NO:Y, a polypeptide encoded by SEQ ID NO:X, and/or a polypeptide encoded by the cDNA in the related cDNA clone contained in a deposited library. Polynucleotides encoding a polypeptide comprising, or alternatively consisting of, the polypeptide sequence of SEQ ID NO:Y, a polypeptide encoded by SEQ ID NO:X, and/or a polypeptide encoded by the the dDNA in the related cDNA clone contained in a deposited library, are also encompassed by the invention. The present invention further encompasses a polynucleotide comprising, or alternatively consisting of, the complement of the nucleic acid sequence of SEQ ID NO:X, and/or the complement of the coding strand of the related cDNA clone contained in a deposited library.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence would unduly burden the disclosure of this application. Accordingly, for each "Contig Id" listed in the first column of Table 3, preferably excluded are one or more polynucleotides comprising a nucleotide sequence described in the second column of Table 3 by the general formula of a-b, each of which are uniquely defined for the SEQ ID NO:X corresponding to that Contig Id in Table 1. Additionally, specific embodiments are directed to polynucleotide sequences excluding at least one, two, three, four, five, ten, or more of the specific polynucleotide sequences referenced by the Genbank Accession No. for each Contig Id which may be

included in column 3 of Table 3. In no way is this listing meant to encompass all of the sequences which may be excluded by the general formula, it is just a representative example.

His Ser Gly Arg Lys Lys
225 230

<210> 1378

<211> 75

<212> PRT

<213> Homo sapiens

<400> 1378

Gly Lys Gln Lys Pro Leu Ser Ser Ala Phe His Leu Gln Glu Arg Arg
1 5 10 15

Lys Asn Ser Cys Leu Leu Ser Val Ile Gln Phe Ala Cys Ile Leu Cys
20 25 30

Ser Cys Thr Asn Pro Tyr Arg Val Asn Leu Leu Ser Thr Ile Tyr Trp
35 40 45

Cys Leu Ile Glu Asn Asp Cys Leu Pro Ser Phe Leu Val Pro Phe Leu
50 55 60

Thr Val Leu Lys Tyr Leu Lys Cys Ile Asp Cys
65 70 75

<210> 1379

<211> 239

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (229)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (231)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (234)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 1379

Arg Arg Gly Gln Val Gly Ala Arg Ser Cys Cys Phe Trp Phe Ser Cys
1 5 10 15

Gly Arg Arg Arg Cys Pro Ala Ala Leu Gly Cys Arg Thr Asp Lys Ala
 20 25 30
 Trp Ala Thr Ala Pro Gln Lys Pro Thr Gln Leu Asp Ala Gly Ala Gly
 35 40 45
 Arg Arg Val Gly Asp Arg Val Ser Glu Gly Ala Ala Arg Ala Gly Gly
 50 55 60
 Arg Ala Pro Glu Gly Glu Arg Gly Gly Gly Gly Ser Ala Ala Gly
 65 70 75 80
 Arg Ala Gly Arg Gly Met Ser Met Pro Asp Ala Met Pro Leu Pro Gly
 85 90 95
 Val Gly Glu Glu Leu Lys Gln Ala Lys Glu Ile Glu Asp Ala Glu Lys
 100 105 110
 Tyr Ser Phe Met Ala Thr Val Thr Lys Ala Pro Lys Lys Gln Ile Gln
 115 120 125
 Phe Ala Asp Asp Met Gln Glu Phe Thr Lys Phe Pro Thr Lys Thr Gly
 130 135 140
 Arg Arg Ser Leu Ser Arg Ser Ile Ser Gln Ser Ser Thr Asp Ser Tyr
 145 150 155 160
 Ser Ser Ala Ala Ser Tyr Thr Asp Ser Ser Asp Asp Glu Val Ser Pro
 165 170 175
 Arg Glu Lys Gln Gln Thr Asn Ser Lys Gly Ser Ser Asn Phe Cys Val
 180 185 190
 Lys Asn Ile Lys Gln Ala Glu Phe Gly Arg Arg Glu Ile Glu Ile Ala
 195 200 205
 Glu Gln Asp Met Ser Ala Leu Ile Ser Leu Arg Lys Arg Ala Gln Gly
 210 215 220
 Glu Lys Pro Leu Xaa Gly Xaa Lys Ile Xaa Gly Leu Thr His Tyr
 225 230 235

<210> 1380

<211> 97

<212> PRT

<213> Homo sapiens

<400> 1380